

**POLIMORFISMOS DE SEQÜÊNCIA NUCLEOTÍDICA EM  
FRAGMENTOS GENÔMICOS DE CANA-DE-AÇÚCAR  
HOMÓLOGOS A GENES DE RESISTÊNCIA**

**MARIANA SENNA QUIRINO**

Dissertação apresentada à Escola Superior de  
Agricultura "Luiz de Queiroz", Universidade de  
São Paulo, para obtenção do título de Mestre  
em Agronomia, Área de Concentração:  
Fitopatologia.

**PIRACICABA**  
Estado de São Paulo - Brasil  
Abril – 2003

**POLIMORFISMOS DE SEQÜÊNCIA NUCLEOTÍDICA EM  
FRAGMENTOS GENÔMICOS DE CANA-DE-AÇÚCAR  
HOMÓLOGOS A GENES DE RESISTÊNCIA**

**MARIANA SENNA QUIRINO**

Engenheira Agrônoma

Orientador: Prof. Dr. **LUIS EDUARDO ARANHA CAMARGO**

Dissertação apresentada à Escola Superior de  
Agricultura "Luiz de Queiroz", Universidade de  
São Paulo, para obtenção do título de Mestre  
em Agronomia, Área de Concentração:  
Fitopatologia.

**P I R A C I C A B A**

Estado de São Paulo – Brasil

Abril - 2003

**Dados Internacionais de Catalogação na Publicação (CIP)  
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Quirino, Mariana Senna

Polimorfismos de seqüência nucleotídica em fragmentos genômicos de  
cana-de-açúcar homólogos a genes de resistência / Mariana Senna  
Quirino. - - Piracicaba, 2003.

52 p.

Dissertação (mestrado) - - Escola Superior de Agricultura Luiz de Queiroz,  
2003.

Bibliografia.

1. Cana-de-açúcar 2. Escapaldadura da folha 3. Ferrugem (Doença de  
planta) 4. Genes 5. Nucleotídeos 6. Polimorfismo 7. Resistência genética  
vegetal 8. Variedades vegetais I. Título

CDD 633.61

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte –

O autor”

*Nossas vidas tocam algumas vidas apenas uma vez nesta caminhada; e depois os caminhos se separam para nunca mais se encontrarem. Portanto não chore porque terminou, mas sorria porque aconteceu. Siga tranqüilamente, entre a inquietude e a pressa, lembrando-se que há sempre paz no silêncio. Quando virmos em Jesus os nossos caminho e paz, Deus colocará beleza e cor nas pequenas coisas que nos acontecem, pois tudo sucede por alguma razão. Então mantenha boas relações com todas as pessoas, amando o seu próximo como a si mesmo, falando a sua verdade, mansa e claramente, e ouça a dos outros, mesmo a dos insensatos e ignorantes, pois também eles têm sua própria história. Portanto, somos agora mais fortes e prontos para continuar em nossa caminhada.*

***"Há tempo de plantar, e tempo de arrancar o que se plantou" (Ec 3.2).***

Ao meu marido José de Souza Marques Neto,  
pelo amor, compreensão e carinho  
**DEDICO.**

Em memória da minha mãe Helena Senna Quirino  
A meu pai Enéas Nunes Quirino  
a todos os meus irmãos e a Tia Dora  
**OFEREÇO.**

## **AGRADECIMENTOS**

Em primeiro lugar a Deus pelo amparo nos momentos difíceis e pela vitória alcançada;

Aos meus sogros Patrício P. Gomes (em memória) e Elza de S. M. P. Gomes pelo apoio e a compreensão;

Ao Professor Luis Eduardo Aranha Camargo pela orientação e oportunidade;

À Professora e amiga Regina de Melo Sartori Coelho pelo incentivo;

Às Doutoradas Cláudia Vitorello e Maeli Melloto pela atenção dispensada;

A todos os professores do Departamento de Entomologia, Fitopatologia e Zoologia Agrícola pela formação acadêmica;

Aos amigos Alice; Kátia; Adriane; Rodrigo; Aninha; Maria Teresa; Célia; Daniela; Jaqueline; Viviane; pela troca de conhecimento e pelos bons momentos;

À querida amiga Marília por tudo que passamos;

Aos companheiros e amigos de Laboratório Paulo; Alê; Cris; Carol; Camila;

Flávia; Giovana; Osmar; pelos momentos agradáveis que passamos;

Aos colegas do curso e aos funcionários do Setor de Fitopatologia;

À Escola Superior de Agricultura "Luiz de Queiroz";

Ao Departamento de Entomologia, Fitopatologia e Zoologia Agrícola;

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP);

À Cooperativa de produtores de cana-de-açúcar e álcool do Estado de São Paulo (Copersucar);

A todos que ajudaram de alguma forma, meus sinceros agradecimentos.

## SUMÁRIO

	Página
LISTA DE FIGURAS.....	viii
LISTA DE TABELAS.....	x
RESUMO.....	xi
SUMMARY.....	xiii
1 INTRODUÇÃO.....	1
2 REVISÃO DE LITERATURA.....	3
2.1 A cultura da cana-de-açúcar.....	3
2.2 Homologia de seqüências conservadas em genes de resistência.....	4
2.3 Projeto SUCEST.....	8
2.4 Polimorfismos de seqüência nucleotídica (SNPs).....	9
3 MATERIAL E MÉTODOS.....	12
3.1 Seleção dos ESTs.....	12
3.2 Material vegetal.....	12

3.3 Amplificação e clonagem de fragmentos genômicos homólogos a genes de resistência.....	14
3.4 Seqüenciamento e análise.....	17
3.5 Cortes com enzima de restrição.....	18
3.6 Construção de iniciadores a partir dos SNPs.....	19
4 RESULTADOS.....	21
4.1 Seleção de ESTs.....	21
4.2 Amplificação e clonagem de fragmentos genômicos homólogos aos genes <i>Xa1</i> e <i>Rp1-D</i> .....	24
4.3 Análise das seqüências dos fragmentos genômicos amplificados com os iniciadores XA1-1F e XA1-1R.....	25
4.3.1 Confirmação das seqüências por meio de enzimas de restrição.....	31
4.3.2 Síntese de iniciadores seqüência-específicos.....	32
4.4 Análise das seqüências dos fragmentos genômicos amplificados com os iniciadores RP1B.F e RP1B.R.....	34
5 DISCUSSÃO.....	39
6 CONCLUSÕES.....	44
REFERÊNCIAS BIBLIOGRÁFICAS.....	45



## LISTA DE FIGURAS

	Página
1 Representação esquemática do alinhamento das seqüências do gene <i>Xa1</i> (barra sólida) e do EST SCCCCL3080E03.g (barra tracejada). Linhas anguladas representam íntrons.....	23
2 Representação esquemática do alinhamento das seqüências do gene <i>Rp1-D</i> (barra sólida) e do EST SCCCCL3080B03.g (barra tracejada). Linha angulada representa íntron.....	23
3 Eletroforese em gel de agarose dos produtos amplificados por PCR com os iniciadores: XA1.1F/XA1.1R a partir do DNA genômico das variedades SP804966 e SP80180 (1 e 2 ) e com os iniciadores RP1B.F/RP1B.R a partir do DNA genômico das variedades R570 e SP811763 (3 e 4). M corresponde ao padrão de peso molecular de 100 pb.....	24
4 Alinhamento de seqüências reproduzíveis de nucleotídeos de fragmentos homólogos ao EST SCCCCL3080E03.g e ao gene <i>Xa1</i> identificadas nas variedades SP804966 e SP80180. Iniciadores SCRP1, SCRP2 e SCRP3 estão indicados em azul, verde e amarelo, respectivamente.....	28
5 Alinhamento de seqüências reproduzíveis de aminoácidos de fragmentos homólogos ao EST SCCCCL3080E03.g e ao gene <i>Xa1</i> identificadas nas variedades SP804966 e SP80180.....	30
6 Eletroforese em gel de agarose das clivagens dos alelos. Os números 1, 2, 3 e 4 correspondem a amostras sem digestão com peso molecular de 379 pb. Os números 1a, 2a, 3a e 4a correspondem aos alelos digeridos	

- com a endonuclease *AluI* gerando fragmentos de 76 pb e 303 pb. Os números 1p, 2p, 3p e 4p correspondem aos alelos digeridos com a endonuclease *PstI* gerando fragmentos de 108 pb e 271 pb. Os números 1e, 2e, 3e e 4e correspondem aos alelos digeridos com a endonuclease *EcoRV* gerando fragmentos de 183 pb e 196 pb. A letra M corresponde ao padrão de peso molecular de 100 pb (Promega)..... 32
- 7 Otimização da PCR com o kit (PCR optimizer™ kit, Invitrogen™), usando tampão com o pH 9,0 e 1,5 mM Mg<sup>2+</sup> (E1 e E2); 2,5 mM Mg<sup>2+</sup> (G1 e G2); 3,5 mM Mg<sup>2+</sup> (H1 e H2); tampão com o pH 9,5 e 1,5 mM Mg<sup>2+</sup> (I1 e I2); 2,5 mM Mg<sup>2+</sup> (K1 e K2); 3,5 mM Mg<sup>2+</sup> (L1 e L2). Para as amostras E1, G1, H1, I1, K1, L1 o DNA genômico pertence à variedade SP804966; para E2, G2, H2, I2, K2, L2 o DNA genômicos pertence à variedade SP80180. M corresponde ao padrão de peso molecular de 1Kb (Promega)..... 33
- 8 Otimização da PCR com o kit (PCR optimizer™ kit, Invitrogen™), usando os tampões com o pH de 8,5 com a concentração de 1,5 mM, para as amostras A1 (DNA genômico da variedade SP804966); A2 (DNA genômico das variedade SP80180); e com o pH 10,0, com concentração de 2,5 mM Mg<sup>2+</sup>; para as amostras B1 (DNA genômico da variedade SP804966) e B2 (DNA genômico da variedade SP80180). A variante foi temperatura de hibridização: 54 °C e 57 °C. M corresponde ao padrão de peso molecular de 1Kb (Promega)..... 34
- 9 Alinhamento de seqüências reproduzíveis de nucleotídeos de fragmentos homólogos ao EST SCCCCL3080E03.g e ao gene *Rp1-D* identificadas nas variedades R570 e SP811763..... 37
- 10 Alinhamento de seqüências reproduzíveis de aminoácidos de fragmentos homólogos ao EST SCCCCL3080E03.g e ao gene *Rp1-D* identificadas nas variedades R570 e SP811763..... 38

## LISTA DE TABELAS

		Página
1	Genes de resistência clonados (adaptada de Hulbert et al., 2001).....	5
2	Níveis de resistência a <i>X. albilineans</i> das cultivares utilizadas neste estudo.....	13
3	Níveis de resistência a <i>P. melanocephala</i> das cultivares utilizadas neste estudo.....	14
4	Nível de homologia de seqüências de nucleotídeos (valor e) entre ESTs de cana-de-açúcar e genes de resistência de diversos hospedeiros que operam em sistemas gene-a-gene.....	22
5	Posições de SNPs (pb) entre seis grupos de seqüências da variedade SP80180 amplificados com o par de iniciadores XA1-1F/XA1-1R.....	26
6	Posições de SNPs (pb) entre quatro grupos de seqüências reproduzíveis da variedade SP804966 amplificados com o par de iniciadores XA1-1F/XA1-1R.....	27
7	Número de SNPs entre seqüências das variedades SP80180 e SP804966 em um fragmento de 379 pb do loco homólogo ao gene <i>Xa1</i> .....	29
8	Sítios de restrição presentes nas seqüências reproduzíveis.....	31
9	Posições de SNPs entre seis seqüências reproduzíveis da variedade SP811763 amplificados com o par de iniciadores RP1-B.F/R.....	35
10	Posições de SNPs entre quatro seqüências reproduzíveis da variedade R570 amplificados com o par de iniciadores RP1-B.F/R.....	36
11	Número de SNPs entre alelos das variedades SP811763 e R570 em fragmentos de 489 pb do loco homólogo ao gene <i>Rp1-D</i> .....	38

# **POLIMORFISMOS DE SEQÜÊNCIA NUCLEOTÍDICA EM FRAGMENTOS GENÔMICOS DE CANA-DE-AÇÚCAR HOMÓLOGOS A GENES DE RESISTÊNCIA**

Autora: MARIANA SENNA QUIRINO

Orientador: Prof. Dr. LUIS EDUARDO ARANHA CAMARGO

## **RESUMO**

Este trabalho objetivou investigar a existência de polimorfismos de seqüência nucleotídica (SNPs "single nucleotide polymorphisms") em fragmentos genômicos de cana-de-açúcar homólogos a genes de resistência. Para tal, iniciadores sintéticos homólogos às extremidades de uma seqüência expressa identificada (ESTs "expressed sequence tag") como homóloga ao gene *Xa1* de arroz (EST SCCCCL3080E03.g) e outra ao gene *Rp1-D* de milho (EST SCCCCL3080B03.g) foram utilizados para amplificar, em duplicata, fragmentos genômicos de variedades de cana-de-açúcar. Em seguida, os fragmentos foram clonados e seqüenciados. No caso do primeiro EST, foram obtidas seqüências de 119 insertos de uma variedade resistente (SP804966) e 156 de uma variedade suscetível (SP80180) a *Xanthomonas albilineans*. No segundo caso, foram obtidas seqüências de 167 insertos de uma terceira variedade resistente (R570) e 135 de outra suscetível (SP811763) a *Puccinia melanocephala*. Para cada EST considerado, as seqüências foram comparadas por meio do programa DNA Sequencher 3.0 (Gene Codes Corporation, Ann

Arbor, MI). Nesta análise, somente foram consideradas seqüências reproduzíveis, isto é, que ocorreram nas duas repetições. No caso de ambos ESTs, a comparação de seqüências entre variedades possibilitou a identificação de quatro a seis fragmentos distintos. A comparação entre variedades, por sua vez, revelou a existência de até duas seqüências comuns. Presume-se que estas seqüências reproduzíveis e confirmadas por meio de digestões de clones representativos de cada uma com enzimas de restrição, correspondam a seqüências alélicas. Iniciadores foram sintetizados com base em polimorfismos encontrados entre os diferentes alelos. No entanto, tentativas de amplificar alelos específicos por meio destes iniciadores se revelaram infrutíferas. Tal insucesso deve-se possivelmente à condição polialélica da cana-de-açúcar, que faz com que polimorfismos entre dois alelos sejam compensados por monomorfismos entre estes e os demais alelos. Assim, a utilização de SNPs como marcadores moleculares baseados em PCR em cana-de-açúcar mostra-se mais complexa quando comparada a plantas diplóides.

# **SINGLE NUCLEOTIDE POLYMORPHISMS IN GENOMIC FRAGMENTS OF SUGARCANE HOMOLOGOUS TO RESISTANCE GENES**

Author: MARIANA SENNA QUIRINO

Adviser: Prof. Dr. LUIS EDUARDO ARANHA CAMARGO

## **SUMMARY**

The objective of this work was to investigate the existence of single nucleotide polymorphisms (SNPs) in genomic fragments homologous to resistance genes in sugarcane. For this purpose, primers were designed based on the sequence of the extremities of an identified expressed sequence tag (ESTs) similar to the *Xa1* gene of rice (EST SCCCCL3080E03.g) and another to the maize *Rp1-D* gene (EST SCCCCL3080B03.g). These primers were used to amplify genomic fragments of sugarcane varieties in duplicate. The fragments were then cloned and sequenced. In the case of first EST, sequences of 119 inserts from a resistant variety (SP804966) and 156 from a susceptible variety (SP80180) to *Xanthomonas albilineans* were analyzed. In the case of the second EST, sequences of 167 inserts from a third variety (R570) and 135 of a fourth one (SP811763), resistant and susceptible respectively to *Puccinia melanocephala*, were analyzed. Sequences were compared using the program DNA Sequencher 3.0 (Gene Codes Corporation, Ann Arbor, MI). In this analysis,

only reproducible sequences were considered, that is, sequences that occurred in both replicates. Four to six different sequences were identified within varieties whereas comparisons among varieties revealed the existence of up to two sequences in common. These reproducible sequences, which were further confirmed through digestion of representative clones with appropriate restriction enzymes, could correspond to allelic sequences. Primers were designed based on the SNPs detected among these so-called alleles. However, attempts to amplify specific alleles using these primers were unsuccessful. This could be due to the polyallelic condition of sugarcane in which polymorphisms between two alleles could be compensated by monomorphisms between these and other alleles. Thus the use of SNPs as PCR-based molecular markers is not as straightforward in sugarcane as in diploid species.

## 1 INTRODUÇÃO

Variedades comerciais de cana-de-açúcar são oriundas de cruzamentos interespecíficos de *Saccharum officinarum* e *Saccharum spontaneum* seguidos de retrocruzamentos usando *Saccharum officinarum* como espécie genitora recorrente, podendo possuir apenas 5 a 10% do genoma das espécies parentais doadoras (Lu et al., 1994). O resultado é um germoplasma com elevado nível de ploidia e de complexo comportamento meiótico, sendo talvez por isto, dentre os membros da família Poacea, a espécie menos caracterizada do ponto de vista genético. Com o advento do projeto SUCEST ("Sugarcane Expressed Sequence Tag") de seqüenciamento de ESTs ("expressed sequence tag"), este panorama está se alterando, já que o projeto possibilitou a identificação de inúmeros genes. Dentre estes, perto de uma centena de ESTs mostraram-se similares a genes de resistência a patógenos. Estes agora representam uma valiosa fonte de informação que pode ser utilizada em estudos de mapeamento de genes de resistência.

Polimorfismos na seqüência nucleotídica de genes homólogos a genes de resistência foram identificados em trabalhos preliminares com os ESTs SCCCCL3080E03.g e SCCCCL3080B03.g. Estes ESTs apresentam similaridade, respectivamente, com o gene *Xa1* que confere resistência a *Xanthomonas oryzae* em arroz (Yoshimura et al., 1998), com o gene e *Rp1-D* de milho, que confere resistência a *Puccinia sorghi* (Hulbert, 1997). Como *Xanthomonas* e *Puccinia* também são patógenos de cana-de-açúcar, e valendo-se da notável conservação genômica existente entre os membros da família



*Poacea* (Grivet et al., 1994; Pereira & Lee, 1995; Moore et al., 1995; Paterson et al., 1996), estes ESTs foram eleitos para análises mais detalhadas descritas neste trabalho, com vistas a identificar polimorfismos de seqüência nucleotídica (SNPs) entre fragmentos genômicos de cana-de-açúcar e avaliar a possibilidade de utilizá-los no desenvolvimento de marcadores moleculares. Marcadores SNPs são relatados como de grande potencial para distinguir alelos em espécies vegetais (Howard et al., 1999). Em milho, por exemplo, estima-se uma freqüência de um polimorfismo a cada 70 bases (Bhatramakki et al., 2000). Já em trigo, Wolters et al. (2000) relataram a diversidade nucleotídica em locos homólogos a genes envolvidos na síntese de amido e constataram a existência de três haplótipos distintos para cada gene analisado que, segundo os autores, representam os três genomas ancestrais que compõem o genoma do trigo cultivado. No caso de cana-de-açúcar, vale lembrar que relatos sobre SNPs em genes homólogos a genes de resistência são inexistentes na literatura.

## 2 REVISÃO DE LITERATURA

### 2.1 A cultura da cana-de-açúcar

A cana-de-açúcar é uma das mais importantes culturas agrícolas do Brasil, tanto pela área que ocupa em regiões tropicais e temperadas, como pelo complexo industrial que gira em torno de seu cultivo, gerando empregos e contribuindo para a alimentação humana e animal. Produzida em 130 países, no Brasil é cultivada em mais de quatro bilhões de hectares e destinada, principalmente, à produção de açúcar e de álcool, à fabricação de aguardente e à alimentação de bovinos. O Brasil é líder na produção e exportação de derivados de cana-de-açúcar, bem como na utilização desta planta como fonte de energia líquida renovável (Matsuoka et al., 1999). Considerando a atual crise energética brasileira, a cultura é também relatada como promissora na produção da energia elétrica a ser extraída do bagaço da cana (Oliveira & Vasconcelos, 2001).

A cana-de-açúcar é uma planta alógama, da família *Gramineae* ou *Poaceae*, tribo *Andropogoneae*, gênero *Saccharum*. No gênero *Saccharum* ocorrem seis espécies: *Saccharum officinarum* L. (2n=80), *S. robustum* Brandes e Jeswit (2n=60-205), *S. barberi* Jeswit (2n=81-124), *S. sinense* Roxb (2n=111-120), *S. spontaneum* L. (2n=40-128) e *S. edule* Hassk (2n=60-80) (Daniels et al., 1987). As cultivares modernas são híbridos oriundos de cruzamentos entre *S. officinarum* e *S. spontaneum* seguidos de retrocruzamentos usando *S. officinarum* como espécie parental recorrente. O resultado é um germoplasma

com elevado nível de ploidia, de complexo comportamento meiótico e que pode possuir até 5 a 10% do genoma das espécies parentais doadoras (Lu et al., 1994). O trabalho de Silva et al. (1993), pode ser citado como um dos pioneiros na área de caracterização genética de cana-de-açúcar e ilustra a dificuldade em se trabalhar com cana-de-açúcar, uma vez que os autores tiveram que recorrer a materiais haplóides e dihaplóides para simplificar a análise genética em populações segregantes. Este trabalho também forneceu os primeiros indícios da extensa colinearidade genômica existente entre cana-de-açúcar e outros membros da família *Poacea*. Embora exista um considerável número de variedades e clones com boas características agrícolas, limitações como suscetibilidade à doenças, alta porcentagem de floração e outras características indesejadas, tornam tais variedades obsoletas. Isso porque a correção desses problemas através de retrocruzamentos é extremamente difícil para a cultura. Em vista disso, a introdução de genes de resistência à doença em novas variedades por meios alternativos, como o mapeamento molecular e transgenia, pode ser uma alternativa de considerável importância nos programas de melhoramento (Silva et al., 1993).

## **2.2 Homologia de seqüências conservadas em genes de resistência**

O advento das técnicas da biologia molecular tornou possível a clonagem e a caracterização de genes de resistência. Os primeiros genes clonados foram *Hm1* em milho (Johal & Briggs, 1992), que codifica uma redutase dependente de NADPH que inativa uma potente proteína produzida pelo patógeno *Cochliobolus carbonum*, e *Pto* em tomate (Martin et al., 1993), que codifica uma proteína similar a proteínas serina-treonina quinases e que confere resistência a *Pseudomonas syringae*. A partir de então, já foram identificados e caracterizados muitos genes em várias espécies vegetais (Tabela 1).

Tabela 1. Genes de resistência clonados (adaptada de Hulbert et al., 2001).

Classe/Estrutura	Gene	Espécie vegetal	Patógeno
TIR:NBS:LRR	<i>L</i>	linho	<i>Melampsora lini</i>
TIR:NBS:LRR	<i>M</i>	linho	<i>Melampsora lini</i>
TIR:NBS:LRR	<i>N</i>	fumo	Vírus TMV
TIR-NBS-LRR	<i>P</i>	linho	<i>Melampsora lini</i>
TIR-NBS-LRR	<i>RPP1</i>	arabidopsis	<i>Peronospora parasitica</i>
TIR-NBS-LRR	<i>RPP5</i>	arabidopsis	<i>Peronospora parasitica</i>
TIR-NBS-LRR	<i>RPP4</i>	arabidopsis	<i>P. syringae pv. tomato</i>
NBS-LRR	<i>RX1</i>	batata	Vírus PVX
NBS-LRR	<i>RPM1</i>	arabidopsis	<i>P. syringae pv. maculicola</i>
NBS-LRR	<i>Mi</i>	tomate	<i>M. incognita</i>
NBS-LRR	<i>RX2</i>	batata	Vírus PVX
NBS-LRR	<i>Bs2</i>	Pimentão	<i>Xanthomonas</i>
NBS-LRR	<i>Dm3</i>	alface	<i>Bremia</i>
NBS-LRR	<i>Prf<sup>b</sup></i>	tomate	<i>Pseudomonas</i>
NBS-LRR	<i>Mla</i>	cevada	<i>B. graminis</i>
NBS-LRR	<i>Rpp8/HRT</i>	arabidopsis	Peronospora; TCV (HRT)
NBS-LRR	<i>RPP13</i>	arabidopsis	<i>Peronospora</i>
NBS-LRR	<i>RPS2</i>	arabidopsis	<i>Pseudomonas</i>
NBS-LRR	<i>RPS5</i>	arabidopsis	<i>Pseudomonas</i>
NBS-LRR	<i>Rp1-D</i>	milho	<i>Puccinia sorghi</i>
NBS-LRR	<i>Xa-1</i>	arroz	<i>X. oryzae pv. oryzae</i>
NBS-LRR	<i>I2C</i>	tomate	<i>F. oxy. F.sp. lycopersici</i>
NBS-LRR	<i>SW5</i>	tomate	Vírus TSWV
NBS-LRR	<i>Pib</i>	arroz	<i>Magnaporthe grisea</i>
NBS-LRR	<i>Pi-ta</i>	arroz	<i>Magnaporthe grisea</i>
TM-LRR	<i>Cf-4</i>	tomate	<i>Cladosporium fulvum</i>
TM-LRR	<i>Cf-5</i>	tomate	<i>Cladosporium fulvum</i>
Pto (prot. quinase)	<i>Pto</i>	tomate	<i>P. syringae pv. tomato</i>
LRRquinase	<i>Xa21</i>	arroz	<i>X. oryzae pv. oryzae</i>
-	HS1 <sup>pro-1</sup>	beterraba	<i>Heterodera schachtii</i>
AS:CC	<i>Rpw8</i>	arabidopsis	<i>E. chichoracearum</i>
TM	<i>mlo</i>	cevada	<i>B. graminis f. sp. hordei;</i>

LRR = regiões ricas em leucina; NBS = sítio aderência nucleotídicos; TIR = "toll/interleukin-1-receptor";

TM = domínio transmembrana; Quinase = proteína; AS = âncora –sinal; CC = espiral enrolado.

A caracterização molecular dos genes de resistência revelou a existência de domínios conservados. Entre os domínios mais comuns encontrados estão: zíper de leucina (LZ), sítio de adesão a nucleotídios (NBS), repetições ricas em leucinas (LRR) e domínios com função de quinases. Para cada um destes domínios foi proposta uma função. Os domínios LZ, por exemplo, supostamente facilitam a união e interação entre duas proteínas, uma vinda do sinal do patógeno e outra de reconhecimento do hospedeiro. Já os domínios LRR

seriam regiões de ligação com as moléculas elicitoras, produtos de genes de avirulência do patógeno. Os domínios NBS, por sua vez, podem possuir atividade de ligação a ATP ou GTP, cujo grupo fosfato seria transferido pelos domínios quinases aos componentes de uma cascata de sinais, que culminaria com a resposta de defesa por parte da planta (Bent, 1996; Hammond-Kosack & Jones 1997).

A existência de tais domínios conservados pode ser utilizada na estratégia de clonagem de genes homólogos a genes de resistência em espécies pouco caracterizadas do ponto de vista genético (Lewin, 2001). Os primeiros trabalhos que relataram o uso de domínios conservados para tal utilizaram iniciadores baseados nas seqüências conservadas dos genes *N*, *RPS2* e *L6* para amplificar regiões genômicas em soja. Os produtos amplificados foram comparados entre si por hibridização e nove classes distintas foram encontradas. Clones representativos destas classes apresentaram homologia de seqüências de aminoácidos de até 66% com os genes utilizados para a elaboração dos iniciadores. Membros destas classes foram utilizados como marcadores em uma população segregante mapeando genes em oito grupos de ligação, sendo que alguns deles mostraram-se ligados aos genes *Rps1* e *RPS2* de resistência à podridão de raiz causada por *Phytophthora soja*, ao gene *Rmd* de resistência a oídio (*Microsphaeria diffusa*), e a um loco de resistência quantitativa para o nematóide do cisto da soja (*Heterodera glycines* Ichinohe). Em outro trabalho, também em soja, fragmentos homólogos a genes de resistência mostraram-se ligados aos genes *Rsv1* (resistência ao vírus do mosaico comum da soja), *Rpv* (resistência ao vírus do mosqueado do amendoim), *Rps1*, *RPS2* e *Rps3* (resistência a *Phytophthora soja*) e *Rmd* (resistência a oídio) (Kanazin et al., 1996; Yu et al., 1998). Em batata, esta mesma estratégia foi utilizada para mapear os gene *Gro1* (resistência a *Globodera rostochiensis*) e *R7* de resistência a *Phytophthora infestans*. Os fragmentos amplificados apresentaram de 43 a 82 % de

homologia em nível de seqüência nucleotídica com os genes *Rps2*, *N* e *L6* (Leister et al., 1996).

Em milho, por sua vez, foram identificadas 11 classes de seqüências baseadas em regiões conservadas dos genes *L6*, *N*, *Rpm1*, *Rpps2* e *12C-1*. Essas seqüências foram utilizadas como sondas e algumas foram mapeadas em regiões contendo locos de resistência já conhecidos, como o gene *Ht* de resistência a *Exserohilum turcicum* e alguns locos de resistência à ferrugem comum (*Puccinia sorghi*) (Collins et al., 1998).

No caso de cana-de-açúcar vários ESTs homólogos a genes R foram identificados pelo projeto SUCEST, entre estes de especial interesse, os clones SCCCCL3080E03.g e SCCCCL3080B03.g, que apresentam homologia com o gene *Xa1* (acesso ao GenBank nº AB002266) de arroz, que confere resistência a *Xanthomonas oryzae*, e *Rp1-D* de milho (acesso ao GenBank nº AF107293), que confere resistência a *Puccinia sorghi*. Como *Xanthomonas* e *Puccinia* também são parasitas de cana-de-açúcar e valendo-se da notável conservação genômica existente entre os membros da família Poacea (Grivet et al., 1994; Pereira & Lee, 1995; Moore et al., 1995; Paterson et al., 1996), estes dois ESTs são interessantes do ponto de vista de sua utilização como marcadores candidatos em trabalhos de mapeamento de genes de resistência a *Xanthomonas albilineans* e *Puccinia melanocephala*.

O gene *Xa1* que confere resistência em arroz à raça 1 de *Xanthomonas oryzae* pv. *oryzae* foi isolado pela estratégia de clonagem baseada em mapa genético e a seqüência de sua proteína mostrou regiões com motivos repetidos de leucina (LRR) e sítio de adesão de nucleotídeo (NBS). É interessante ressaltar que o gene *Xa1* difere dos demais genes de resistência porque tem expressão constitutiva; somente é expresso sob condições de infecção ou injúria mecânica (Yoshimura et al., 1998). O gene *Rp1* é um complexo loco de resistência à ferrugem de milho. O haplótipo *Hrp1-D* é composto de *Rp1-D* e oito parálogos, sete dos quais codificam proteínas com os domínios LRR-NBS, semelhantes ao gene *Rp1-D*. Esses parálogos são polimórficos especialmente

nas regiões da proteína que correspondem aos motivos repetidos do aminoácido leucina. O último membro da família (*Hrp1-D*) codifica uma proteína que contém domínio LRR. A comparação das seqüências entre os parálogos sugere que a origem desta série de alelos ocorreu a partir de eventos de recombinação, principalmente nos domínios LRR, com diversificação no processo de seleção. Alterações do gene *Rp1-D*, que resultaram em perda completa ou parcial da resistência, foram explicadas pela possibilidade da ocorrência de “crossing over” desigual em regiões codificadoras (Collins et al., 1999; Hulbert 1997; Sun et al., 2001).

### 2.3 Projeto SUCEST

Existem basicamente dois tipos de projetos genoma. Um baseado no seqüenciamento total do genoma e outro baseado apenas no seqüenciamento de genes expressos. A estratégia mais utilizada no primeiro caso é a chamada “shotgun”, que consiste no seqüenciamento em grande escala de pequenos fragmentos genômicos até a completa organização do genoma. O seqüenciamento total, no entanto, se aplica preferencialmente a genomas pequenos, o que não é o caso de cana-de-açúcar. Considerando esse aspecto, o seqüenciamento de seqüências expressas identificadas (EST ou “Expressed Sequenced Tag”) é uma alternativa menos custosa e que apresenta ainda a vantagem de caracterizar a expressão temporal e local dos genes (Carneiro et al., 2000).

O Projeto “Sugarcane Expressed Sequence Tag” SUCEST - <http://sucest.lad.ic.uni.camp.br>), realizado pelo consórcio “Organization for Nucleotide Sequencing and Analysis” (ONSA - <http://watson.fapesp.br>) da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), seqüenciou 292 mil ESTs e identificou mais de 43 mil genes de cana-de-açúcar (Vettore et al., 2001). Para a espécie *Saccharum officinarum*, pode-se antecipar

um impacto positivo na compreensão do seu complexo genoma e, conseqüentemente, no melhoramento genético da espécie. Analisando os ESTs, por exemplo, pode-se inferir sobre fatores biológicos vitais ao desenvolvimento da planta, como vias de transduções de sinais, genes responsáveis pelo desenvolvimento da planta, estresse biótico e abiótico, regulação gênica, presença de elementos transponíveis, análise de polimorfismos, constituição das organelas e vários outros assuntos relacionados ao metabolismo da planta (Arruda, 2001).

A comparação de seqüências pode ser feita com base nos nucleotídeos, aminoácidos ou domínios funcionais. As análises são feitas comparando as seqüências obtidas com aquelas editadas em banco de dados depositadas na Internet e os resultados são avaliados por meio de um valor probabilístico (“e value”). Quanto menor o valor deste índice, maior a possibilidade da seqüência comparada ser similar àquela do banco de dados. A partir dos dados de seqüenciamento e utilizando algoritmos de alinhamento como ferramenta, foi identificada uma centena de ESTs com homologia a genes de resistência que representam uma valiosa fonte de marcadores candidatos que podem ser usados em trabalhos de mapeamento genômico (Carneiro et al., 2000; <http://sucest.lad.ic.uni.camp.br>).

## **2.4 Polimorfismos de seqüência nucleotídica (SNPs)**

Nos últimos anos, o rápido desenvolvimento de técnicas para marcadores moleculares contribuiu significativamente para a elucidação de questões relativas à genética de populações e de indivíduos. Os primeiros marcadores descritos eram relativamente complexos e dependiam de técnicas de “Southern Blot” para a sua detecção (RFLPs). Recentemente, foram descritos marcadores de detecção mais simples, baseados na técnica de PCR (Santos, 1995).



Polimorfismos de seqüência nucleotídica (SNP) representam uma fonte abundante de variação genética. Esses polimorfismos são gerados por substituição de uma única base ou pequenos eventos de inserção e deleção, ocorrendo numa taxa muito baixa, de aproximadamente  $10^{-7}$  alterações por sítios por geração, ou menos (Hammer et al., 1997). A grande vantagem dos SNPs em comparação a outros marcadores reside na abundância de polimorfismos entre alelos de um determinado gene. Em humanos, por exemplo, SNPs ocorrem a uma freqüência de 1 a cada 1000 nucleotídeos (Nickerson et al., 1998; Pennisi, 1998; Collins et al., 1999). Em vegetais, informações acerca desse tipo de marcadores são escassos e possivelmente uma freqüência maior possa ser encontrada. No caso de *Arabidopsis thaliana*, estimativas apontam para a existência de 56.670 polimorfismos (The Arabidopsis Genome Initiative, 2000). Bhatramakki et al., (2000) apresentaram trabalho referente à freqüência de SNPs em milho e relataram a ocorrência de um polimorfismo a cada 70 bases, em média. Destes, 60% foram transições e 40% transversões. Os autores detectaram ainda uma deleção/inserção a cada 160 pb. Ao examinarem o padrão de ocorrência de SNPs, os autores puderam definir haplótipos e observaram pouca diversidade entre eles para genótipos de procedências distintas em fragmentos de até 300 pb, uma provável conseqüência da intensa domesticação aplicada a esta espécie. Ao analisarem fragmentos maiores, os autores encontraram evidências de recombinação entre haplótipos. O objetivo dos autores foi o de estabelecer um catálogo de SNPs em mais de 1000 locos selecionados de ESTs e genes de interesse comercial. Já Wolters et al. (2000) relataram a diversidade nucleotídica em locos homólogos de trigo. Através da amplificação, clonagem e seqüenciamento de fragmentos de genes envolvidos na síntese de amido, os autores constataram a existência de três haplótipos distintos para cada gene analisado, os quais representam os três genomas ancestrais que compõem o genoma do trigo cultivado. O objetivo do trabalho de Wolters et al. (2000) foi o de desenhar iniciadores alelo-específicos de cada genoma, valendo-se de SNPs, e avaliar a

diversidade dos mesmos entre as diversas linhagens de trigo na tentativa de correlacionar esta diversidade com variabilidades funcionais ou fenotípicas existentes entre as linhagens.

SNPs também foram usados para diferenciar alelos de um gene de resistência à antracnose em feijoeiro. Melloto & Kelly (2001) estudaram o polimorfismo de seqüência nucleotídica em uma população segregante para resistência a antracnose de feijoeiro conferida pelo o gene *CO4-K*. No alinhamento das seqüências dos genitores, os SNPs presentes foram observados na forma de substituições, inserções e deleções. Comparando homólogos de *CO4-K* em diferentes genótipos, os SNPs foram mais acurados em discriminar genótipos de feijoeiro quanto à presença do loco *CO4-K* do que um marcador “Sequenced Characterized Amplified Region” - SCAR (SAS13) proximamente ligado a esse gene. Estes SNPs cosegregaram com o fenótipo resistência e mostraram-se ligados a três alelos diferentes no loco *CO4-K*.

Soori et al. (1999), estudaram a freqüência de SNP em batata. Um fragmento de DNA do hospedeiro ligado ao gene *Ry<sub>adg</sub>* de resistência ao vírus PVY, foi amplificado por PCR e seqüenciado. Verificou-se que a substituição de um nucleotídeo aboliu o único local de reconhecimento da enzima de restrição (*Bbv* I), resultando na substituição do aminoácido nessa posição (Ser → Thr) somente para o genótipo de batata que carregava o gene *Ry<sub>adg</sub>*. Os autores consideraram que o SNP pode ser determinante de um gene de resistência ou parte do gene *Ry<sub>adg</sub>*.

## **3 MATERIAL E MÉTODOS**

### **3.1 Seleção dos ESTs**

Para uma seleção preliminar de ESTs, foram realizadas buscas no banco de dados do projeto SUCEST (<http://sucest.lad.ic.unicamp.br>) usando palavras-chave relacionadas à resistência a doença (“disease resistance”, “NBS-LRR”, “resistance like gene”, etc) e os parâmetros valor limite (“threshold value”) de –30 a 0, penalidade (“penalty”) de -2, concordância mínima (“minimum match”) de 16 e escore mínimo (“minimum score”) de 30. Numa segunda seleção, foram selecionados dois ESTs homólogos a genes de resistência de gramíneas e com pouca homologia entre si.

### **3.2 Material vegetal**

O material vegetal utilizado neste trabalho foi cedido pela Copersucar e escolhido com base na seleção dos ESTs. Assim, no caso do primeiro EST, foram utilizadas as cultivares SP80180 e SP804966, respectivamente suscetível e resistente a *Xanthomonas albilineans*, bem como clones F<sub>1</sub> originados do cruzamento dessas variedades que foram previamente avaliados quanto a resistência a este patógeno (A. Sanguino, dados não publicados. Tabela 2).

No caso do segundo EST, foram utilizadas cultivares com níveis variados de resistência e suscetibilidade à ferrugem segundo avaliação prévia (A. Sanguino, dados não publicados. Tabela 3).

Tabela 2. Níveis de resistência a *X. albilineans* das cultivares utilizadas neste estudo

<b>Variedades</b>	<b>severidade<sup>1</sup></b>	<b>Clones<sup>2</sup></b>	<b>severidade<sup>1</sup></b>
SP921197	9	234-105	9
SP931323	2	234-22,	8
SP901963	8	234-25	8
SP813250	7	234-119	9
SP823349	9	234-133	8
SP913616	1	234-129	9
SP913617	1	234-201	9
SP913650	1	234-138	4
SP913660	1	234-179	8
SP913665	1	234-132	8
SP913667	1	234-185	-
SP913702	8	234-414	-
SP924140	-		
SP784767	9		
SP821044	8		
SP785495	8		
SP913636	1		
SP913661	1		

<sup>1</sup> Notas segundo escala diagramática onde 1 significa– ausência da doença e 9 altamente suscetível; Clones<sup>2</sup> SP80180X SP804966.

Tabela 3. Níveis de resistência a *P. melanocephala* das cultivares utilizadas neste estudo

Variedades	severidade <sup>1</sup>	Variedades	severidade <sup>1</sup>
R570	2	SP931099	1
SP811763	7	SP931074	1
SP931334	1	SP931088	1
SP931411	8	SP931092	1
SP933247	8	SP891117	1
SP931138	1	SP891117M	7
SP931202	1	SP891038	9
SP931210	8	SP88673	3
SP931262	1	SP791011	9
SP931275	8	SP701143	9
SP931332	1	RB72454	2

<sup>1</sup> Notas segundo escala diagramática onde 1 significa ausência da doença e 9 altamente suscetível

### 3.3 Amplificação e clonagem de fragmentos genômicos homólogos a genes de resistência

Iniciadores homólogos às extremidades dos ESTs selecionados foram utilizados para amplificar fragmentos genômicos homólogos a genes de resistência. As amplificações por reação de polimerase em cadeia (PCR) foram realizadas em dois ensaios utilizando 30 ng/μL de DNA depositados em microtubos contendo tampão de amplificação 1X, 0,2 mM de deoxinucleotídeos trifosfatos (dNTPs) (Life Technologies, Rockville, MD), 0,25 μM de cada iniciador e 2,5 U de *Taq* DNA Polymerase (Promega, Madison, WI, EUA) ou 0,3 U de Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA – USA) em um volume final de 20 μL. As amplificações foram feitas em um termociclador PTC 100 (MJ Reserarch, EUA) utilizando um ciclo a 93 °C por 2 minutos seguido de 35 ciclos de desnaturação a 93°C por 45 segundos, de

hibridização a “X” °C por 45 segundos e de extensão a 72 °C por 1 minuto, seguidos de um ciclo final de extensão por 10 minutos a 72 °C. A temperatura “X” de hibridização foi calculada para cada par de iniciadores segundo a fórmula  $T_m = [2\text{ }^\circ\text{C (A + T) + 4 }^\circ\text{C (C + G) - 5 }^\circ\text{C}]$  (Newton & Graham, 1997).

O produto da amplificação foi resolvido em gel de agarose 1%. A corrida foi feita a 8 V/cm a 37 °C. Logo após, o gel foi corado com 0,5  $\mu\text{g mL}^{-1}$  de brometo de etídeo, visualizado e fotodocumentado com aparelho Imagem Master® VDS (Pharmacia Biotech, San Francisco, EUA). Para a verificação do tamanho do fragmento, foi utilizado o marcador 1kb DNA Ladder (Promega Madison, WI, EUA).

Após a amplificação, os fragmentos foram submetidos à ligação em vetor plasmidial com o kit pGEM®-T Easy Vector Systems (Promega Madison, WI, EUA), segundo as instruções do fabricante. As amostras no volume final de 10  $\mu\text{L}$  foram incubadas a 4 °C por 10 a 16 horas e armazenadas a -20 °C para posterior transformação.

Para a transformação, foram utilizadas células competentes de *Escherichia coli* estirpe DH5 $\alpha$  com a eficiência de  $10^7$  UFCmL<sup>-1</sup> preparadas segundo Hanahan (1983). Aos 200  $\mu\text{L}$  de células competentes acondicionados em tubos de 1,5 mL, foram adicionados de 2 a 8  $\mu\text{L}$  do produto da ligação. Essa mistura foi incubada por 30 minutos no gelo, submetida a choque térmico por incubação a 42 °C por 90 segundos e imediatamente transferida para o gelo. Em ambiente asséptico, a cada tubo foram adicionados 800  $\mu\text{L}$  de meio SOC (1% de triptona; 0,5% extrato de levedura; 8,5% NaCl 2,5 mM KCl; 0,01 mM MgCl<sub>2</sub>; 0,02 mM de glicose, água q.s.p. 1000 mL). A mistura foi incubada a 37 °C por 1 h sob agitação orbital de 220 rpm. Em seguida, as células foram centrifugadas a 12054 x g (rotor F1802, centrífuga Microfuge ® Lite centrifuge, Palo Alto, CA- USA) por 1 minuto e 400  $\mu\text{L}$  do sobrenadante resultante foram descartados e as células suspensas no volume restante com auxílio de pipeta. Aproximadamente 100  $\mu\text{L}$  dessa suspensão foram distribuídos em

placas de petri contendo meio LB sólido adicionado de 0,1 mM IPTG(isopropil- $\beta$ -D-tiogalactopiranosida), 40  $\mu$ g de X-gal (5-bromo-4cloro-3indolil- $\beta$ -D-galactopiranosida) e de 100  $\mu$ g mL<sup>-1</sup> de ampicilina. As placas de Petri foram incubadas a 37 °C por 12-14 horas.

Para extração do DNA plasmidial, as colônias recombinantes foram transferidas para 1 mL de meio Circle Grow (Bio 101 Inc., Carlsbad, CA) acrescido de 100  $\mu$ g mL<sup>-1</sup> de ampicilina em placas de crescimento de 96 amostras (Masterblock, Greiner Lab., Frickenhausen, Al.) e cultivadas sob agitação orbital a 320 rpm por 22 horas a 37 °C. As microplacas foram então centrifugadas por 6 minutos a 514,20 x g em centrífuga JOUAN, modelo BR4i (Jouan, SA., Saint Herblain, Fr) com rotor JOUAN Swing-out S20 para sedimentar as células. O sobrenadante foi descartado e o sedimentado bacteriano foi suspenso em 240  $\mu$ L de GET (Glicose 20%; EDTA 0,5 M pH 8,0; Tris-HCl 1 M pH 7,4) por meio de agitação por 2 minutos. A seguir, as células foram novamente sedimentadas por centrifugação de 8 minutos a 514,20 x g. O sobrenadante foi descartado e a placa invertida sobre papel absorvente por 5 minutos. Foram adicionados mais 80  $\mu$ L de GET e a placa agitada por 2 minutos. Aproximadamente 60  $\mu$ L de cada amostra foram transferidos para placas de polipropileno de fundo arredondado (Millipore, SA., Molshein, Fr.) contendo 2,5  $\mu$ L de RNase (10 mg mL<sup>-1</sup>). Nas amostras foram adicionados 80  $\mu$ L de uma solução contendo NaOH 0,2N e SDS 1%, sendo as mesmas invertidas 30 vezes e incubadas por 10 minutos em temperatura ambiente. A cada célula da placa foram adicionados 80  $\mu$ L de KOAc 3M, previamente resfriado a 4 °C. As placas foram então novamente invertidas por 30 vezes, incubadas por mais 10 minutos em temperatura ambiente seguidos de 30 minutos em estufa a 90 °C. A seguir, foram resfriadas em gelo e centrifugadas por 8 minutos a 514,20 x g a 5 °C. O sobrenadante foi transferido para placa "Millipore" modelo MAGV N22 (Millipore, SA., Molshein, Fr) fixada sobre uma microplaca de polipropileno de fundo tipo "V" (Corning Inc., Corning, NY, USA) e

centrifugado por 5 minutos a 514,20 x g a 5 °C. A placa Millipore foi descartada e adicionados 100 µL de isopropanol (Merck), sendo as placas invertidas por 30 vezes e centrifugadas por 45 minutos a 514,20 x g a 5 °C. O sobrenadante foi descartado, sendo adicionado ao sedimentado 200 µL de Etanol 70% (Merck) gelado. Centrifugou-se por 5 minutos a 514,20 x g a 5 °C e removeu-se o sobrenadante. Inverteu-se a placa sobre papel absorvente e centrifugou-se a 115,71 g e 20 °C por 30 segundos. O sedimento foi seco em temperatura ambiente por 60 minutos e, logo em seguida, o DNA foi suspenso em 40 µL de água estéril. As amostras foram incubadas em temperatura ambiente por 12 horas e armazenadas a -20 °C.

O seqüenciamento dos insertos dos clones foi realizado pelo método Sanger (1977) com os kits Big Dye TM "Terminator Cycle Sequencing Ready Reaction" (PE Applied Biosystems, Foster City CA) e Dyenamic™ ET "Terminator Cycle Sequencing Kit" (Armershan Pharmacia Biotech, Madison - USA). Foram utilizados 400 ng de DNA e 3,2 µM do iniciador específico (anverso ou reverso), com volume final de 10 µL. As amostras foram submetidas às seguintes condições de amplificação em termociclador Geneamp PCR System (Perkin Elmer, CA): 96 °C de desnaturação inicial por 1 minuto, 40 ciclos de 96 °C por 20 segundos; 52 °C por 20 segundos; 60 °C por 4 minutos. Os dideoxynucleotídeos (ddNTPs) não incorporados foram precipitados em placas de 96 células, seguindo as recomendações dos fabricantes. Após a precipitação, as amostras foram suspensas em 3 µL de tampão "Blue Dextran" (Promega, Madison, WI, USA).

### **3.4 Seqüenciamento e análise**

O seqüenciamento foi realizado no laboratório de Genética Molecular do Departamento de Entomologia, Fitopatologia e Zoologia Agrícola da



ESALQ/USP. Os nucleotídeos foram resolvidos por eletroforese em gel 5% de poliacrilamida Long Ranger® gel Solution 50% (BioWhittaker Molecular Applications - BMA, Rockland, ME, USA) em aparelho ABI Prism 377 DNA Sequencer™ (PE Applied Biosystems, Foster City, CA, USA) por sete horas, de acordo com o protocolo descrito pelo fabricante.

As seqüências de nucleotídeos dos clones foram alinhadas por meio do programa DNA Sequencher 3.0 (Gene Codes Corporation, Ann Arbor, MI). Discrepâncias entre seqüências foram averiguadas por inspeções visuais dos eletroferogramas e editadas manualmente, quando necessário. Após, este procedimento foram eliminadas seqüências não reproduzíveis, isto é, aquelas que só ocorreram em um dos ensaios. As seqüências reproduzíveis foram alinhadas e comparadas pelo programa Multalin do INRA que permitiu identificar a presença de SNPs. A identidade das seqüências foi confirmada comparando-as às seqüências dos ESTs originais depositadas no banco SUCEST e às seqüências depositadas no “GenBank” através do software “Blast” (Altschul et al., 1997).

### **3.5 Cortes com enzima de restrição**

O programa “DNAClub” (Xiongfong Chen, Universition of Cornell, Ithaca, NY - <http://128.84.203.244>) e o programa *DNA Sequencher 3.0* (Gene Codes Corporation, Ann Arbor, MI) foram utilizados para gerar um mapa de restrição das seqüências. Insetos de clones representativos de cada seqüência foram amplificados com iniciadores posicionados na extremidade 5' do fragmento e clivados com enzimas de restrição segundo o mapa de restrição. A reação de amplificação procedeu-se como descrito no item 3.3. Porém, antes da clivagem, os produtos da PCR foram purificados com o Kit Wizard® PCR Preps DNA Purification System (Promega Corporation - Madison, WI USA). As digestões foram realizadas seguindo o protocolo do fabricante, utilizando 1 µg do produto

da PCR purificado, 10 u/μL da respectiva enzima de restrição e tampão da enzima diluído a 1X em um volume final de 20 μL. A digestão foi realizada a 37°C por 12 horas e o produto foi separado em gel de agarose de 2% (*Amplisize Agarose Bio-Rad Laboratories*) por eletroforese (5V/cm) e visualizado após coloração do mesmo com 0,5 μgmL<sup>-1</sup> de brometo de etídeo e fotodocumentado com aparelho Imagem Master® VDS (Pharmacia Biotech, San Francisco, EUA).

### **3.6 Construção de iniciadores a partir dos SNPs**

Na tentativa de amplificar seqüência-específicos de cada variedade, iniciadores foram construídos com base na análise e posicionamento dos SNPs, por meio do programa “DNAClub” (By Xiongfong Chen, Universition of Cornell, Ithaca, NY - <http://128.84.203.244>). A temperatura de hibridização variou de acordo com o par de iniciadores utilizado. Para SCRP1/SCRP3 esta foi de 57°C, ao passo que para SCRP2/SCRP3 foi de 54°C. A reação da PCR em um volume final de 20 μL, foi composta por 30 ng/μL DNA, 0,2 mM de deoxinucleosídeos trifosfatos (dNTPs) (Life Technologies, Rockeville, MD); 1X tampão de amplificação, 0,25 μM de cada iniciador e 2,5 U de *Taq* DNA Polymerase (Promega, Madison, WI, EUA). As amplificações foram feitas em um termociclador PTC 100 (MJ Reserarch, EUA) utilizando um ciclo a 93 °C por 2 minutos seguidos de 35 ciclos de desnaturação a 93°C por 45 segundos, de anelamento a “X” °C por 45 segundos e de extensão a 72 °C por 1 minuto, seguidos de um ciclo final de extensão por 10 minutos a 72 °C. A temperatura “X” de hibridização foi calculada para cada par de iniciadores. Com as duas combinações de iniciadores foram amplificados os fragmentos genômicos das variedades SP80180 e SP804966 e de quinze plantas F<sub>1</sub>. Não era esperado que a seqüência-específicos de variedades resistentes fossem amplificados

com o iniciador SCRP3, uma vez que este foi construído a partir dos SNPs dos indivíduos suscetíveis.

Outra tentativa para a amplificação das seqüência-específicos foi realizada utilizando o kit PCR optimizer™, Invitrogen™, Carlsbad, Califórnia - USA. Para a otimização das condições de amplificação foram realizados cinco testes, sendo nestes casos testados pH do tampão, concentração de Mg<sup>2+</sup>; presença ou ausência de DMSO, DNA polimerase e temperatura de hibridização. Nestes testes foi utilizado DNA genômico das variedades SP80180 e SP804966 na concentração de 30 ng/μL, 0,2 mM de deoxinucleosídeos trifosfatos (dNTPs) e, 0,25 μM de cada iniciador em um volume final de 20 μL. Deste modo, o arranjo experimental ficou do seguintes modo: no teste 1 foi utilizado tampão com o pH 8,5 e MgCl<sub>2</sub> variando nas concentrações de 1,5; 2,0; 2,5 e 3,5 mM ; tampão com o pH 9,0 e 2,0 mM MgCl<sub>2</sub> ; tampão com o pH 9,5 e 2,0 mM Mg<sup>2+</sup>; tampão com o pH 10,0 e 2,0 mM MgCl<sub>2</sub>. Neste teste a temperatura de hibridização foi de 54 °C para todas as combinações; no teste 2 foi utilizado o tampão com o pH de 9,0 e MgCl<sub>2</sub> variando as concentrações de 1,5; 2,5 e 3,5mM; pH 9,5 e variando as concentrações de MgCl<sub>2</sub>: 1,5; 2,5,e 3,5 mM ; pH 10,0 e MgCl<sub>2</sub> variando as concentrações de 1,5; 2,5,e 3,5 mM. Neste teste a temperatura de hibridização foi de 54 °C para todas as combinações; no teste 3 foi usado DMSO 10% usando o tampão com o pH de 8,5 e 1,5, 2,5 mM Mg<sup>2+</sup> e pH de 10,0 e 1,5, 2,5 mM Mg<sup>2+</sup>; pois foi feita nestas condições a amplificação dos fragmentos desejados com temperatura de hibridização de 54 °C; no teste 4 as mesmas condições de amplificação do teste 3, variando apenas a temperatura de hibridização para 57 °C e por último o teste 5 as mesmas condições de amplificação do teste 3, variando apenas a DNA polimerase: 2,5 U de *Taq* DNA Polymerase (Promega, Madison, WI, EUA) e 0,3 U de Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA – USA).

## 4 RESULTADOS

### 4.1 Seleção de ESTs

Buscas realizadas no banco de dados do projeto SUCEST identificaram vários ESTs homólogos a genes de resistência que operam dentro do sistema clássico gene-a-gene de Flor. Dentre estes, seis ESTs apresentaram homologia significativa com genes de resistência de várias espécies hospedeiras (Tabela 4). Alguns deles apresentaram homologia a mais de um gene de resistência, o que não é de todo inesperado uma vez que genes de resistência apresentam domínios conservados entre si. Em adição, vários outros ESTs apresentaram homologia a putativos genes de resistência. São denominados putativos, pois têm função desconhecida, mas apresentam domínios característicos de genes de resistência.

Os ESTs escolhidos para o desenvolvimento deste trabalho foram SCCCCL3080E03.g e SCCCCL3080B03.g, obtidos de biblioteca de calo da variedade SP80-3280 e submetidos a estresse de temperatura (4°C e 37°C) e luz (alternância de claro / escuro). Esses clones foram selecionados por apresentarem homologia com os genes *Xa1* de arroz e *Rp1-D* de milho. A razão desta escolha se deve ao fato da proximidade genética entre cana-de-açúcar e estes hospedeiros e também porque os gêneros *Puccinia* e *Xanthomonas* abrigam espécies patogênicas a *Saccharum*.

O primeiro EST possui 589 pares de bases homólogas à região que possui dois domínios conservados e um motivo NBS do gene *Xa1* (Figura1). O

segundo EST, de 685 pares de base, homólogo à extremidade 3' do gene *Rp1-D* (Figura 2), que corresponde ao domínio NBS e LRRs. Ambos ESTs não possuem íntrons na região homólogas aos genes.

Tabela 4. Nível de homologia de seqüências de nucleotídeos (valor *e*) entre ESTs de cana-de-açúcar e genes de resistência de diversos hospedeiros que operam em sistemas gene-a-gene.

EST	Gene de resistência	hospedeiro	patógeno	valor <i>e</i>
SCCCCL3080B03.g	<i>Rp1-D</i>	milho	<i>Puccinia sorghi</i>	e 00
SCCCCL3080E03.g	<i>Xa-1</i>	arroz	<i>Xanthomonas o. pv. oryzae</i>	8e-30
	<i>I2C-1</i>	tomate	<i>Fusarium</i>	2e-17
	<i>RPM1</i>	<i>Arabidopsis</i>	<i>Pseudomonas syringae pv maculicola</i>	8e-11
SCCCCL4017A12.g	<i>Pto</i>	tomate	<i>Pseudomonas syringae pv. tomato</i>	2e-53
SCCCLR1072H06.g	<i>Pto</i>	tomate	( <i>Pseudomonas syringae pv. tomato</i> )	2e-63
SCCCLR1C04F01.g	RX	batata	vírus X (PVX)	3e-28
	<i>RPM1</i>	<i>Arabidopsis</i>	<i>Pseudomonas syringae pv maculicola</i>	2e-24
	<i>Rpp8</i>	<i>Arabidopsis</i>	<i>Peronospora parasitica</i>	5e-22
	<i>RPS2</i>	<i>Arabidopsis</i>	<i>Pseudomonas syringae</i>	1e-14
	<i>Xa1</i>	arroz	<i>Xanthomonas o. pv. oryzae</i>	6e-12
SCCCRZ1001F12.g	<i>Cf-2.2</i>	tomate	<i>Cladosporium fulvum</i>	3e-38
	<i>Xa21</i>	arroz	<i>Xanthomonas o. pv. oryzae</i>	1e-31

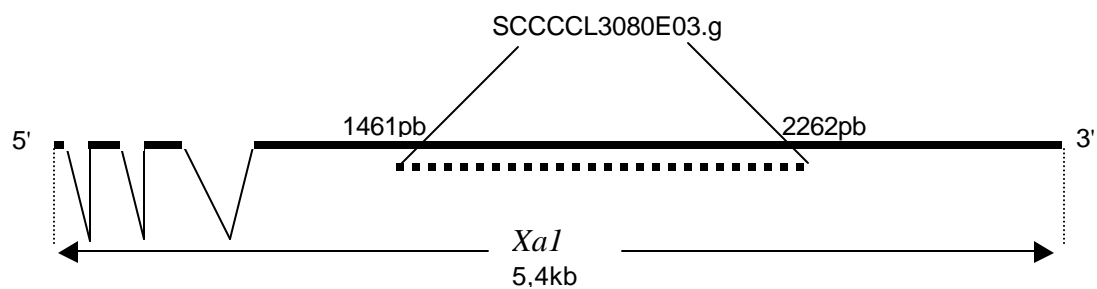


Figura 1 - Representação esquemática do alinhamento das seqüências do gene *Xa1* (barra sólida) e do EST SCCCCL3080E03.g (barra tracejada). Linhas anguladas representam íntrons.

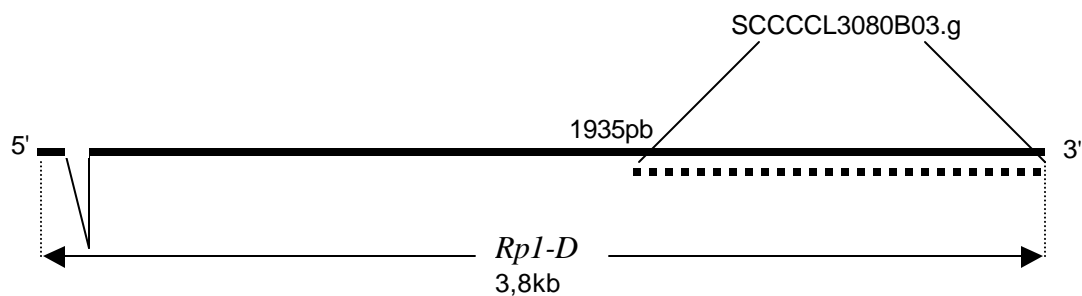


Figura 2 - Representação esquemática do alinhamento das seqüências do gene *Rp1-D* (barra sólida) e do EST SCCCCL3080B03.g (barra tracejada). Linha angulada representa íntron.

## 4.2 Amplificação e clonagem de fragmentos genômicos homólogos aos genes *Xa1* e *Rp1-D*

As seqüências das extremidades do EST SCCCCL3080E03.g foram utilizadas para sintetizar os iniciadores XA1-1F (5'-GGG CAG CAG ATT GTT AAA GCA CTA AAG GG-3') e XA1-1R (5'-GGG ATG TCC CTA GAG TTG GAT CCA-3') e amplificar fragmentos genômicos das variedades SP80180 e SP804966. Em ambas variedades, a PCR amplificou um fragmento único de tamanho aproximado de 680 pb (Figura 3). No caso do segundo EST, SCCCCL3080B03.g, os iniciadores utilizados foram RP1B.F (5'-TCT GAA GCA GTT GAT GGC ACT-3') e RP1B.R (5'-GTG GTT GAG CAA TAC GGA GCT-3') que também resultaram na amplificação de um único fragmento genômico das variedades R570 e SP811763 de tamanho aproximado de 800 pb (Figura 3).

Na segunda amplificação com a mesma combinação de iniciadores e variedades (repetição 2), os resultados foram os mesmos que os observados na primeira amplificação.



Figura 3 - Eletroforese em gel de agarose dos produtos amplificados por PCR com os iniciadores: XA1.1F/XA1.1R a partir do DNA genômico das variedades SP804966 e SP80180 (1 e 2) e com os iniciadores RP1B.F/RP1B.R a partir do DNA genômico das variedades R570 e SP811763 (3 e 4). M corresponde ao padrão de peso molecular de 100 pb.

### **4.3 Análise das seqüências dos fragmentos genômicos amplificados com os iniciadores XA1-1F e XA1-1R**

Na análise das seqüências dos fragmentos genômicos amplificados nas variedades SP80180 e SP804966 com os iniciadores XA1-1F/XA1-1R, somente foram consideradas 379 bases centrais do fragmento. Esta porção foi escolhida por apresentar satisfatória resolução de bases na maioria das leituras a julgar por inspeção visual dos cromatogramas.

Para a variedade SP80180, 156 leituras, escolhidas de clones obtidos das duas repetições, foram utilizadas para análise. O alinhamento das seqüências destes fragmentos utilizando o critério de 100% de homologia e cobertura mínima de 99% resultou no agrupamento de 65 leituras em seis grupos distintos de seqüências reproduzíveis, ou seja, seqüências que ocorreram em ambas repetições. O número de clones nestes agrupamentos de seqüências, no entanto, não foi igual. Um grupo teve a maioria (38 seqüências) reproduzidas nos 65 insertos, enquanto que os outros abrangeram 7,5,7,3 e 5 insertos de seqüências reproduzidas cada. O restante das leituras (91) não foi incluído na análise, pois estas somente ocorreram em uma das repetições (leituras não reproduzíveis).

A comparação das seis seqüências reproduzíveis revelou a existência de 10 SNPs (Tabela 5), indicando uma freqüência de um polimorfismo a cada 37 pares de bases nesta variedade.



Tabela 5. Posições de SNPs (pb) entre seis grupos de seqüências da variedade SP80180 amplificados com o par de iniciadores XA1-1F/XA1-1R.

	Seqüências		Posição do nucleotídeo							
	7	10	148	163	251	332	333	335	342	352
1/18	C	G	<b>T</b>	A	<b>T</b>	<b>G</b>	T	<b>T</b>	G	T
2/18	<b>T</b>	G	C	A	G	A	T	G	G	T
3/18	C	G	C	<b>C</b>	G	A	<b>A</b>	G	G	T
4/18	C	G	C	<b>C</b>	G	A	T	G	<b>A</b>	<b>C</b>
5/18	C	<b>A</b>	<b>T</b>	A	<b>T</b>	<b>G</b>	T	<b>T</b>	G	T
6/18	C	G	C	A	G	A	T	G	G	T

No caso da variedade SP804966, foram seqüenciados 119 insertos sendo que seu alinhamento, utilizando os mesmos critérios, resultou no agrupamento de 62 leituras em quatro grupos distintos de seqüências idênticas e que repetiram nos dois ensaios. Um grupo teve a maioria, 25 seqüências reproduzidas dos 62 insertos, ao passo que os outros abrangeram 20, 11, 6 insertos de seqüências em ambos os ensaios. O restante das leituras (57) compreendeu seqüências não-reproduzíveis. A comparação das seqüências reproduzíveis permitiu identificar 5 SNPs (Tabela 6), resultando em um polimorfismo a cada 75 bases, uma freqüência menor que a da variedade SP80180.

Tabela 6. Posições de SNPs (pb) entre quatro grupos de seqüências reproduzíveis da variedade SP804966 amplificados com o par de iniciadores XA1-1F/XA1-1R.

Seqüências	Posição do Nucleotídeo				
	7	163	335	342	352
1/49	<b>T</b>	<b>A</b>	G	G	T
2/49	C	C	G	G	T
3/49	C	<b>A</b>	<b>T</b>	G	T
4/49	C	C	G	<b>A</b>	<b>C</b>

A comparação das seqüências reproduzíveis identificadas nas variedades SP80180 e SP804966 (Figura 4) revelou a existência de duas seqüências presentes em ambas variedades (seqüências 1/49 4/49, 2/18 4/18). O número de SNPs entre seqüências variou de 0 a 8 (Tabela 7).

As seqüências reproduzidas de ambas variedades foram comparadas a seqüências do EST depositadas no banco SUCEST. Os resultados confirmaram a identidade deles com o EST SCCCL3080E03.g a julgar pelo elevado valor “e” (e = 0.0). A homologia encontrada para nucleotídeos foi de 98,9 a 100% com o citado EST. A seqüência 3/49 foi a que apresentou homologia de 100%.

```

2/18 1 TACATATGAGCACTGGAGGACAGTTCAAAACAAATGGAAATCTCTTCAAGAAGATATTGA
2/49 1 TACATACGAGCACTGGAGGACAGTTCAAACAATGGAAATCTCTTCAAGAAGATATTGA
3/18 1 TACATACGAGCACTGGAGGACAGTTCAAACAATGGAAATCTCTTCAAGAAGATATTGA
4/18 1 TACATACGAGCACTGGAGGACAGTTCAAACAATGGAAATCTCTTCAAGAAGATATTGA
6/18 1 TACATACGAGCACTGGAGGACAGTTCAAACAATGGAAATCTCTTCAAGAAGATATTGA
1/49 1 TACATATGAGCACTGGAGGACAGTTCAAAACAAATGGAAATCTCTTCAAGAAGATATTGA
5/18 1 TACATACGACTGGAGAGGACAGTTCAAAACAAATGGAAATCTCTTCAAGAAGATATTGA
3/49 1 TACATACGAGCACTGGAGGACAGTTCAAACAATGGAAATCTCTTCAAGAAGATATTGA
1/18 1 TACATACGAGCACTGGAGGACAGTTCAAACAATGGAAATCTCTTCAAGAAGATATTGA
4/49 1 TACATACGAGCACTGGAGGACAGTTCAAACAATGGAAATCTCTTCAAGAAGATATTGA

2/18 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
2/49 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
3/18 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
4/18 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
6/18 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
1/49 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
5/18 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
3/49 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
1/18 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC
4/49 61 TGGTATTATCTTGAAGCTAAGTTATGATTTTTTACCAGTCCACCTGCAGCACTGTTTCTC

2/18 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGAAATCTCGTCCATGCTTG
2/49 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGCAATCTCGTCCATGCTTG
3/18 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGCAATCTCGTCCATGCTTG
4/18 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGCAATCTCGTCCATGCTTG
6/18 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGAAATCTCGTCCATGCTTG
1/49 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGAAATCTCGTCCATGCTTG
5/18 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGAAATCTCGTCCATGCTTG
3/49 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGAAATCTCGTCCATGCTTG
1/18 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGAAATCTCGTCCATGCTTG
4/49 121 GTACTACTCTCTGTTTCCAGAGGATTACCAATTTGATGGGGGCAATCTCGTCCATGCTTG

2/18 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
2/49 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
3/18 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
4/18 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
6/18 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
1/49 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
5/18 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
3/49 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
1/18 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA
4/49 181 GATATCACAAAATTTTGTGCAGTGCGAAAGACCCCTCCATGAGATTGGAGGAAACAGGGAA

2/18 241 GCAATATTTGGATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
2/49 241 GCAATATTTGGATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
3/18 241 GCAATATTTGGATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
4/18 241 GCAATATTTGGATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
6/18 241 GCAATATTTGGATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
1/49 241 GCAATATTTGGATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
5/18 241 GCAATATTTGTATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
3/49 241 GCAATATTTGGATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
1/18 241 GCAATATTTGTATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT
4/49 241 GCAATATTTGGATAATTTAGTGGATTGGGCTTCTTCCAAAAGGTTGGCTCAGACTATGT

2/18 301 CATGCATGACCTAATGCATGAATTGGCTCATATGGTTTCGCGAAATGATTGTGCCACTGT
2/49 301 CATGCATGACCTAATGCATGAATTGGCTCATATGGTTTCGCGAAATGATTGTGCCACTGT
3/18 301 CATGCATGACCTAATGCATGAATTGGCTCATAAGGTTTCGCGAAATGATTGTGCCACTGT
4/18 301 CATGCATGACCTAATGCATGAATTGGCTCATATGGTTTCGCAAATGATTGTGCCACTGT
6/18 301 CATGCATGACCTAATGCATGAATTGGCTCATATGGTTTCGCGAAATGATTGTGCCACTGT
1/49 301 CATGCATGACCTAATGCATGAATTGGCTCATATGGTTTCGCGAAATGATTGTGCCACTGT
5/18 301 CATGCATGACCTAATGCATGAATTGGCTCATGTCTTTTCGCGAAATGATTGTGCCACTGT
3/49 301 CATGCATGACCTAATGCATGAATTGGCTCATATGTTTTTCGCGAAATGATTGTGCCACTGT
1/18 301 CATGCATGACCTAATGCATGAATTGGCTCATGTCTTTTCGCGAAATGATTGTGCCACTGT
4/49 301 CATGCATGACCTAATGCATGAATTGGCTCATATGGTTTCGCAAATGATTGTGCCACTGT
...

```

Figura 4 - Alinhamento de seqüências reproduzíveis de nucleotídeos de fragmentos homólogos ao EST SCCCL3080E03.g e ao gene *Xa1* identificadas nas variedades SP804966 e SP80180. Iniciadores SCR1, SCR2 e SCR3 estão indicados em azul, verde e amarelo, respectivamente.

Tabela 7. Número de SNPs entre seqüências das variedades SP80180 e SP804966 em um fragmento de 379 pb do loco homólogo ao gene *Xa1*.

Seq	1/18	2/18	3/18	4/18	5/18	6/18	1/49	2/49	3/49	4/49
1/18	-	5	6	7	1	4	5	5	3	7
2/18	-	-	3	4	6	1	0	2	2	4
3/18	-	-	-	3	7	2	3	1	3	3
4/18	-	-	-	-	8	3	4	2	4	0
5/18	-	-	-	-	-	5	6	6	4	8
6/18	-	-	-	-	-	-	1	1	1	3
1/49	-	-	-	-	-	-	-	2	2	4
2/49	-	-	-	-	-	-	-	-	2	2
3/49	-	-	-	-	-	-	-	-	-	4
4/49	-	-	-	-	-	-	-	-	-	-

O alinhamento das seqüências de aminoácidos de cada variedade, realizado com o auxílio do software Clustal W, pode ser visualizado na Figura 5, como destacado em preto. Pode-se notar que em alguns casos os polimorfismos de seqüências verificados em nível de nucleotídeos refletiram em substituições de aminoácidos.

```

4/18 1 YIRALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLPIZWGQSRPCL
4/49 1 YIRALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLPIZWGQSRPCL
2/49 1 YIRALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLPIZWGQSRPCL
3/18 1 YIRALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLPIZWGQSRPCL
2/18 1 YIZALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLPIZWGKSRPCL
1/49 1 YIZALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLPIZWGKSRPCL
6/18 1 YIRALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLPIZWGKSRPCL
3/49 1 YIRALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLPIZWGKSRPCL
1/18 1 YIRALESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLSIZWGKSRPCL
5/18 1 YIRTDLESSKQMEISSRRYZWYYLEAKLZFFTSPPAALFLVLLSVSRGLSIZWGKSRPCL

4/18 61 DITKFCAVRRPFHEIGGNREAI FGZ FSGFGLLPKGWRLRLCHAZPNAZIGSYGFAKZLRHC
4/49 61 DITKFCAVRRPFHEIGGNREAI FGZ FSGFGLLPKGWRLRLCHAZPNAZIGSYGFAKZLRHC
2/49 61 DITKFCAVRRPFHEIGGNREAI FGZ FSGFGLLPKGWRLRLCHAZPNAZIGSYGFAKZLCHC
3/18 61 DITKFCAVRRPFHEIGGNREAI FGZ FSGFGLLPKGWRLRLCHAZPNAZIGSZGFAKZLCHC
2/18 61 DITKFCAVRRPFHEIGGNREAI FGZ FSGFGLLPKGWRLRLCHAZPNAZIGSYGFAKZLCHC
1/49 61 DITKFCAVRRPFHEIGGNREAI FGZ FSGFGLLPKGWRLRLCHAZPNAZIGSYGFAKZLCHC
6/18 61 DITKFCAVRRPFHEIGGNREAI FGZ FSGFGLLPKGWRLRLCHAZPNAZIGSYGFAKZLCHC
3/49 61 DITKFCAVRRPFHEIGGNREAI FGZ FSGFGLLPKGWRLRLCHAZPNAZIGSYVFAKZLCHC
1/18 61 DITKFCAVRRPFHEIGGNREAI FVZ FSGFGLLPKGWRLRLCHAZPNAZIGSCVFAKZLCHC
5/18 61 DITKFCAVRRPFHEIGGNREAI FVZ FSGFGLLPKGWRLRLCHAZPNAZIGSCVFAKZLCHC

4/18 121 KWIQLZ
4/49 121 KWIQLZ
2/49 121 KWIQLZ
3/18 121 KWIQLZ
2/18 121 KWIQLZ
1/49 121 KWIQLZ
6/18 121 KWIQLZ
3/49 121 KWIQLZ
1/18 121 KWIQLZ
5/18 121 KWIQLZ

```

Figura 5 - Alinhamento de seqüências reproduzíveis de aminoácidos de fragmentos homólogos ao EST SCCCCL3080E03.g e ao gene *Xa1* identificadas nas variedades SP804966 e SP80180.

### 4.3.1 Confirmação das seqüências por meio de enzimas de restrição.

A confirmação das seqüências reproduzíveis foi feita pela clivagem com enzimas de restrição de acordo com mapas de restrição preditos com base em cada seqüência (Tabela 8). Clones representantes de cada seqüência foram amplificados com os iniciadores SCRP1 [5'-TACATACGAGCACTG-3'] e SCPR2 [5'-ACAATCATTTTCGCGAAAACAC-3'] e posteriormente digeridos com enzimas apropriadas.

Os padrões de digestão com *AluI*, *PstI* e *EcoRV* podem ser vistos na Figura 6. As três enzimas reconhecem um sítio de clivagem cada uma presente em todas as seqüências reproduzíveis. Nota-se que os padrões de restrição de clones representativos de cada seqüência-específico confirmam aqueles esperados com base no mapa de restrição. No caso de *AluI*, por exemplo, a digestão resultou em dois fragmentos de aproximadamente 300 e 70pb. De maneira análoga, a digestão com *PstI* gerou fragmentos de aproximadamente, 100 e 270pb. No caso da *EcoRV*, os fragmentos digeridos continham pesos moleculares de aproximadamente 180 e 190pb, também confirmando o padrão previsto.

Tabela 8. Sítios de restrição presentes nas seqüências reproduzíveis.

Enzimas	Sítio de Restrição	seqüência	Nº de sítios	Posição de clivagem (pb)
<b><i>AluI</i></b>	<b>AGCT / TCGA</b>	Todas	1	76
<b><i>NdeI</i></b>	<b>CATATG / GTATAC</b>	1/49, 2/18	2	4 e 330
<b><i>NdeI</i></b>	<b>CATATG / GTATAC</b>	2/49, 3/49, 4/49, 4/18, 6/18	1	330
<b><i>EcoRV</i></b>	<b>GATATC / CTATAG</b>	Todos	1	183
<b><i>PstI</i></b>	<b>CTGCAG / GACGTC</b>	Todos	1	108

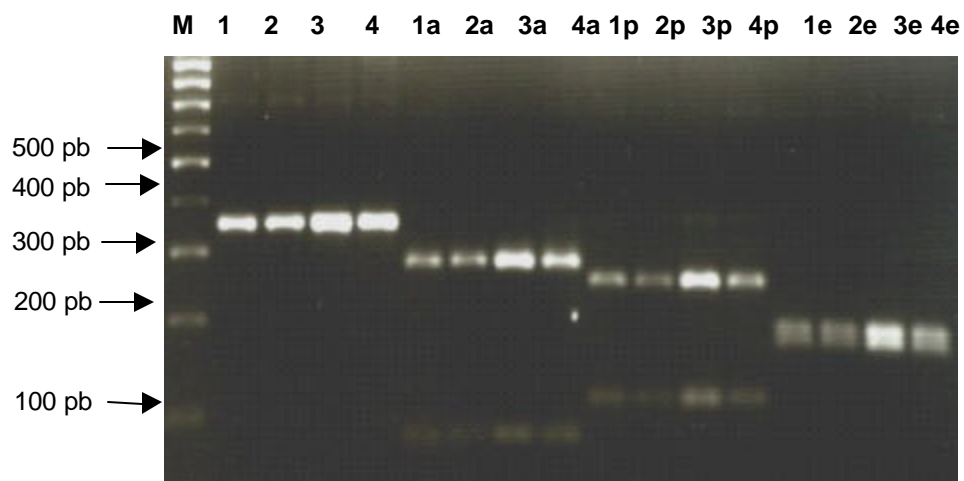


Figura 6 - Eletroforese em gel de agarose das clivagens dos alelos. Os números 1, 2, 3 e 4 correspondem a amostras sem digestão com peso molecular de 379 pb. Os números 1a, 2a, 3a e 4a correspondem aos alelos digeridos com a endonuclease *AluI* gerando fragmentos de 76 pb e 303 pb. Os números 1p, 2p, 3p e 4p correspondem aos alelos digeridos com a endonuclease *PstI* gerando fragmentos de 108 pb e 271 pb. Os números 1e, 2e, 3e e 4e correspondem aos alelos digeridos com a endonuclease *EcoRV* gerando fragmentos de 183 pb e 196 pb. A letra M corresponde ao padrão de peso molecular de 100 pb (Promega).

### 4.3.2 Síntese de iniciadores seqüência-específicos

Os Iniciadores SCRP1 [5'-TACATATGAGCACTG-3', sentido anverso], SCPR2 [5'-TACATACGAGCACTG-3', sentido anverso] e SCPR3 [5'-ACAATCATTTGCGGAAAACAC-3', sentido reverso] foram sintetizados a partir de regiões polimórficas encontradas nas seqüências 1/49, 1/18 e 5/18 na tentativa de obter iniciadores seqüência-específicos (Figura 4).

Outra tentativa para a amplificação das seqüência-específicos foi realizada utilizando o kit (PCR optimizer™ kit, Invitrogen™, Carlsbad, Califórnia - USA). Para a otimização das condições de amplificação foram realizados cinco testes, sendo nestes casos testados pH do tampão, concentração de  $Mg^{2+}$ ; presença ou ausência de DMSO, DNA polimerase e temperatura de

hibridização. Nestes testes foi utilizado DNA genômico das variedades SP80180 e SP804966 como relatado no item 3.6. Resultados de dois testes (2 e 4) estão nas Figuras 7 e 8, respectivamente.

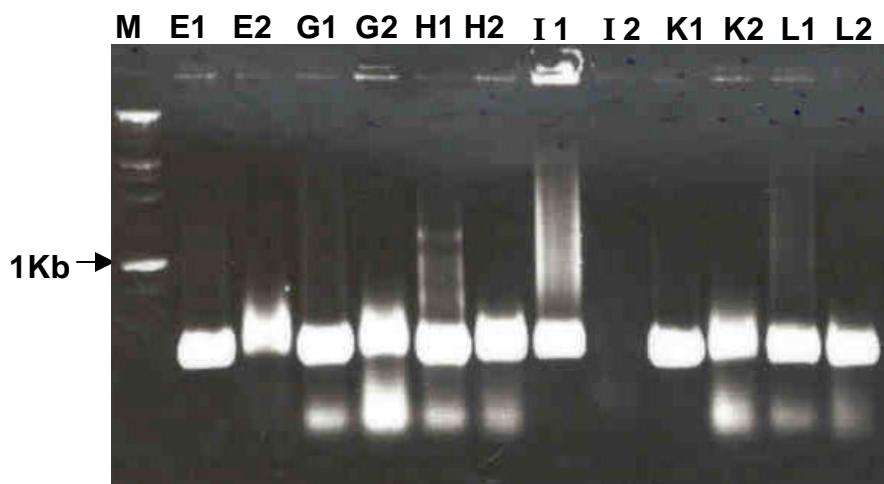


Figura 7 - Otimização da PCR com o kit (PCR optimizer™ kit, Invitrogen™), usando tampão com o pH 9,0 e 1,5 mM Mg<sup>2+</sup> (E1 e E2); 2,5 mM Mg<sup>2+</sup> (G1 e G2); 3,5 mM Mg<sup>2+</sup> (H1 e H2); tampão com o pH 9,5 e 1,5 mM Mg<sup>2+</sup> (I1 e I2); 2,5 mM Mg<sup>2+</sup> (K1 e K2); 3,5 mM Mg<sup>2+</sup> (L1 e L2). Para as amostras E1, G1, H1, I1, K1, L1 o DNA genômico pertence à variedade SP804966; para E2, G2, H2, I2, K2, L2 o DNA genômicos pertence à variedade SP80180. M corresponde ao padrão de peso molecular de 1Kb (Promega).



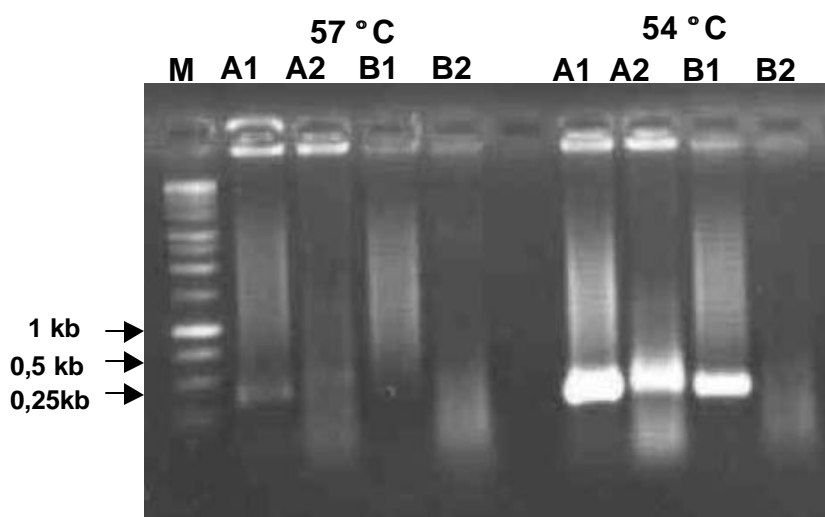


Figura 8 - Otimização da PCR com o kit (PCR optimizer™ kit, Invitrogen™), usando os tampões com o pH de 8,5 com a concentração de 1,5 mM, para as amostras A1 (DNA genômico da variedade SP804966); A2 (DNA genômico das variedade SP80180); e com o pH 10,0, com concentração de 2,5 mM Mg<sup>2+</sup>; para as amostras B1 (DNA genômico da variedade SP804966) e B2 (DNA genômico da variedade SP80180). A variante foi temperatura de hibridização: 54 °C e 57 °C. M corresponde ao padrão de peso molecular de 1Kb (Promega).

#### 4.4 Análise das seqüências dos fragmentos genômicos amplificados com os iniciadores RP1B.F e RP1B.R

Para a variedade SP811763 um total de 135 insertos foi utilizado para análise. Somente foram consideradas as 393 bases centrais do fragmento. Esta porção, a exemplo do fragmento homólogo ao gene *Xa1*, foi escolhida por apresentar satisfatória resolução de bases na maioria das leituras. O alinhamento das seqüências destes fragmentos, utilizando os critérios mencionados anteriormente, resultou no agrupamento de 53 seqüências em seis grupos distintos de seqüências reproduzíveis idênticas entre si. O número de clones nestes agrupamentos, no entanto, não foi igual. Um grupo teve a maioria (17) dos 53 insertos, os outros abrangeram 13, 12, 6, 3 e 2 cada. Ao

todo 82 seqüências não foram incluídas na análise, pois não foram reproduzidas. A comparação das seis reproduzíveis revelou a existência de 28 SNPs (Tabela 9), resultando numa freqüência de 1 polimorfismo a cada 14 pares de bases.

Tabela 9. Posições de SNPs entre seis seqüências reproduzíveis da variedade SP811763 amplificados com o par de iniciadores RP1-B.F / R

Posição do nucleotídeo	Seqüências					
	1/17	2/17	3/17	4/17	5/17	6/17
9	G	G	G	G	G	T
37	C	C	C	C	C	T
48	G	G	<b>A</b>	<b>A</b>	G	G
50	A	A	A	A	<b>C</b>	A
55	<b>A</b>	<b>A</b>	C	C	C	C
121	<b>G</b>	<b>G</b>	C	C	C	C
125	<b>T</b>	<b>T</b>	A	A	A	A
129	G	G	G	<b>A</b>	<b>A</b>	<b>A</b>
130	A	A	A	A	<b>G</b>	<b>G</b>
136	<b>A</b>	<b>A</b>	G	G	G	G
173	A	A	<b>G</b>	<b>G</b>	A	<b>G</b>
188	G	G	G	G	<b>C</b>	G
197	T	T	T	T	T	<b>G</b>
201	A	A	A	A	A	T
202	T	T	G	G	<b>C</b>	G
203	T	T	T	T	T	<b>G</b>
209	A	A	<b>T</b>	<b>T</b>	A	A
250	G	G	G	G	G	<b>T</b>
268	C	C	C	C	<b>T</b>	C
276	T	T	T	T	T	<b>A</b>
277	<b>G</b>	<b>G</b>	A	A	A	A
282	T	T	T	<b>G</b>	T	T
283	A	A	A	<b>G</b>	A	A
286	A	A	A	A	<b>G</b>	A
306	<b>A</b>	<b>A</b>	G	G	G	G
317	A	A	<b>T</b>	A	A	A
348	<b>A</b>	<b>A</b>	G	G	G	G
353	<b>T</b>	<b>T</b>	A	A	A	A

No caso da variedade R570, foram seqüenciados 167 insertos e o alinhamento desses resultou no agrupamento de 106 em quatro grupos de seqüências reproduzíveis. O restante (61) compreendeu seqüências não-reproduzíveis. A comparação das seqüências reproduzíveis permitiu identificar

2 SNPs (Tabela 10), resultando em um polimorfismo a cada 196 bases, isto é, uma frequência menor que a da variedade SP811763.

Tabela 10. Posições de SNPs entre quatro seqüências reproduzíveis da variedade R570 amplificados com o par de iniciadores RP1-B.F/R.

Seqüência	Posição dos Nucleotídeos	
	253	382
1/57	C	T
2/57	C	T
3/57	T	T
4/57	C	C

A comparação entre seqüências reproduzíveis das duas variedades (Figura 9) revelou a existência de duas seqüências presentes em ambas (1/17, 2/17 1/57 e 2/57). O número de SNPs entre seqüências variou de 0 a 21 (Tabela 11). As seqüências reproduzíveis apresentaram homologia de nucleotídeos de 96 a 100% com os ESTs SCUTAD1032F04.g (e = 0.0), e SCCCCL3080B03.g, (e =-118).

```

1/17 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
2/57 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
2/17 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
1/57 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
3/57 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
4/57 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
3/17 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
4/17 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
5/17 1 ATGGCAGGGCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT
6/17 1 ATGGCAGGTCCATGGAGCTGCCAGTGTCTTCTACCATCAGGACTCCGTGAACCTTTACCTTT

1/17 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
2/57 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
2/17 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
1/57 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
3/57 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
4/57 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
3/17 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
4/17 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
5/17 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA
6/17 61 CGTCATGCAGTATTACAGATGAAGCTTTAGCTATTTGCCTTGGTGGCCTCACTTCACTGA

1/17 121 GAACTTTAGAATTGGAATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
2/57 121 GAACTTTAGAATTGGAATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
2/17 121 GAACTTTAGAATTGGAATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
1/57 121 GAACTTTAGAATTGGAATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
3/57 121 GAACTTTAGAATTGGAATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
4/57 121 GAACTTTAGAATTGGAATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
3/17 121 CAACATTAGAATTGGATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
4/17 121 CAACATTAGAATTGGATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
4/17 121 CAACATTAGAATTGGATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
5/17 121 CAACATTAGAATTGGATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC
6/17 121 CAACATTAGAATTGGATACAATATGGCATTAACTGCCTTCCGTGAGAAGAAGTGTTTC

1/17 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
2/57 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
2/17 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
1/57 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
3/57 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
4/57 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
3/17 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
4/17 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
5/17 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG
6/17 181 AGCATGTGACAAAGCTTGAGATTTTCGAAATCACTGCTTGTGGTGTCTCAGATCACTGG

1/17 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTGTTTTATCATTGCCCTTCTTTAG
2/57 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTGTTTTATCATTGCCCTTCTTTAG
2/17 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTGTTTTATCATTGCCCTTCTTTAG
1/57 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTGTTTTATCATTGCCCTTCTTTAG
3/57 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTGTTTTATCATTGCCCTTCTTTAG
4/57 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTGTTTTATCATTGCCCTTCTTTAG
3/17 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTAATTTTTATCATTGCCCTTCTTTAG
4/17 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTAATTTGGTCATTGCCCTTCTTTAG
5/17 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTAATTTTTATCATTGCCCTTCTTTAG
6/17 241 GGGGCTTACGTGCTGCTCCATCTCTTTTCGAGCCTTTAATTTTTATCATTGCCCTTCTTTAG

1/17 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
2/57 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
2/17 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
1/57 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
3/57 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
4/57 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
3/17 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
4/17 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
5/17 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT
6/17 301 AGCTGACACGTGGAGCAGAATTTATGCCGTTGGACCTTAGTTTCTCATCATTAAATGGCT

```

Figura 9 - Alinhamento de seqüências reproduzíveis de nucleotídeos de fragmentos homólogos ao EST SCCCCL3080E03.g e ao gene *Rp1-D* identificadas nas variedades R570 e SP811763.

Tabela 11. Número de SNPs entre alelos das variedades SP811763 e R570 em fragmentos de 489 pb do loco homólogo ao gene *Rp1-D*.

Alelo	1/17	2/17	3/17	4/17	5/17	6/17	1/57	2/57	3/57	4/57
1/17	-	0	13	15	15	19	0	0	1	1
2/17	-	-	13	15	15	19	0	0	1	1
3/17	-	-	-	4	11	13	13	13	14	14
4/17	-	-	-	-	11	13	15	15	16	16
5/17	-	-	-	-	-	14	15	15	15	16
6/17	-	-	-	-	-	-	19	20	21	21
1/57	-	-	-	-	-	-	-	0	1	1
2/57	-	-	-	-	-	-	-	-	1	1
3/57	-	-	-	-	-	-	-	-	-	2
4/57	-	-	-	-	-	-	-	-	-	-

O alinhamento das seqüências de aminoácidos de cada variedade, realizado com o auxílio do software Clustal W pode ser visualizado na Figura 12. Note que em alguns casos os polimorfismos de seqüência verificados em nível de nucleotídeos refletiram em substituições de aminoácidos.

```

3/17      1 QHZNWDTIWHZLHFRQKRCFSMZQSLRVSISLLVGVSDHWGAYVLLHLFRAFVFI IALLZ
4/17      1 QHZNWDTIWHZLHFRQKRCFSMZQSLRVSISLLVGVSDHWGAYVLLHLFRAFVILV IALLZ
1/17      1 ELZNWNTIWHZLHFRQKCCFSMZQSLRFSKSLLVGVSDHWGAYVLLHLFRAFVFI IALLZ
2/57      1 ELZNWNTIWHZLHFRQKCCFSMZQSLRFSKSLLVGVSDHWGAYVLLHLFRAFVFI IALLZ
3/57      1 ELZNWNTIWHZLHFRQKCCFSMZQSLRFSKSLLVGVSDHWGAYVLLHLFRAFVFI IALLZ
4/57      1 ELZNWNTIWHZLHFRQKCCFSMZQSLRFSKSLLVGVSDHWGAYVLLHLFRAFVFI IALLZ
2/17      1 ELZNWNTIWHZLHFRQKCCFSMZQSLRFSKSLLVGVSDHWGAYVLLHLFRAFVFI IALLZ
1/57      1 ELZNWNTIWHZLHFRQKCCFSMZQSLRFSKSLLVGVSDHWGAYVLLHLFRAFVFI IALLZ
5/17      1 QHZDWDTIWHZLHFRQKCCFSMZQSLRLSKSLLVGVSDHWGAYVLLHLFZAFI FIVALLZ
6/17      1 QHZDWDTIWHZLHFRQKRCFSMZQSWSGSKSLLVGVSDHWGAYVLLHLFRALFI IALLZ

3/17      61 SWHVELNLCRWTLVSSSZMAAYLQLIHSLMACRSZRVFTFVV-
4/17      61 SWHVEQNLCRWTLVSSSZMAAYLQLIHSLMACRTZRVFTFMV-
1/17      61 SZHVEQNLCRWTLVSSSLMAAYLQLIHSLMACRTZRMFAFVV-
2/57      61 SZHVEQNLCRWTLVSSSLMAAYLQLIHSLMACRTZRMFAFVV-
3/57      61 SZHVEQNLCRWTLVSSSLMAAYLQLIHSLMACRTZRMFAFVV-
4/57      61 SZHVEQNLCRWTLVSSSLMAAYLQLIHPLMACRTZRMFAFVV-
2/17      61 SZHVEQNLCRWTLVSSSLMAAYLQLIHSLMACRTZRMFAFVV-
1/57      61 SZHVEQNLCRWTLVSSSLMAAYLQLIHSLMACRTZRMFAFVVA
5/17      61 SWHVEQNLCRWTLVSSSZMAAYLQLIHSLMACRTZRIFTFMV-
6/17      61 SWHVEQNLCRWTLVSSSYMAAYLQLIHSLMACRTZKFTFVV-

```

Figura 10 - Alinhamento de seqüências reproduzíveis de aminoácidos de fragmentos homólogos ao EST SCCCL3080E03.g e ao gene *Rp1-D* identificadas nas variedades R570 e SP811763.

## 5 DISCUSSÃO

O Projeto SUCEST destaca-se como pioneiro no sequenciamento em larga escala de genes de uma espécie vegetal. Um de seus resultados foi a criação de um catálogo eletrônico de seqüências gênicas agrupadas de acordo com sua similaridade a genes de outros organismos. Neste catálogo, podem ser encontrados vários genes similares a genes de resistência de diversas espécies vegetais. Um dos objetivos do presente trabalho foi o de avaliar o grau de variabilidade genética em nível de seqüência nucleotídica existente em alguns destes genes. A escolha dos ESTs utilizados neste trabalho deu-se em função de serem muito similares a genes de resistência de espécies gramíneas próximas à cana-de-açúcar e que, além disto, conferem resistência a patógenos (*Xanthomonas* e *Puccinia*) que também são patógenos de *Saccharum*.

Um segundo objetivo foi o de desenvolver uma metodologia de utilização de SNPs encontrados nestes genes como marcadores moleculares baseados na reação da PCR. SNPs são considerados a classe mais informativa de marcadores moleculares, uma vez que são extremamente polimórficos e segregam de maneira co-dominante (Nickerson et al., 1998). Em plantas, já há exemplos da utilização destes marcadores na identificação de variedades, na construção de mapas genéticos e físicos e análises funcionais (Somers et al., 2002).

Os iniciadores desenhados com base nas seqüências dos ESTs escolhidos foram usados para amplificar fragmentos genômicos de linhagens de cana-de-açúcar resistentes e suscetíveis aos patógenos (*Xanthomonas* e *Puccinia*), resultando na amplificação de um único produto de PCR de tamanho

igual ao esperado em todas as linhagens testadas. Trabalho anterior (Camargo 2000), no qual estes fragmentos foram sequenciados diretamente após sua purificação do gel, sugeriu que estes, na verdade, compreendem uma população de fragmentos de igual tamanho, porém diversos em seqüência, que corresponderiam a alelos. Para corroborar esta hipótese, no presente trabalho os fragmentos foram amplificados novamente e os produtos resultantes clonados e sequenciados. Este procedimento foi realizado em duplicata para ambos pares de iniciadores a fim de identificar seqüências reproduzíveis e não-reproduzíveis.

A análise destas seqüências revelou a existência de várias seqüências não reproduzíveis, isto é, que ocorreram uma única vez em apenas uma das duplicatas. Há pelo menos duas interpretações possíveis não excludentes para este fato. Algumas destas seqüências podem ser resultados espúrios da PCR, devidos a erros de substituição de nucleotídeos que normalmente ocorrem durante a polimerização. No entanto, a baixa freqüência deste tipo de erro, estimada entre  $10^{-3}$  a  $10^{-6}$  para polimerases de replicação e reparo (Kunkel et al., 2000) não é suficiente para explicar o grande número de seqüências não reproduzíveis encontradas neste trabalho. Assim, uma segunda explicação poderia ser a de que os iniciadores não são loco-específicos e amplificam alelos de outros locos similares aos dos ESTs estudados. Esta hipótese decorre do fato do genoma da cana conter várias regiões duplicadas (Grivet et al., 1996) e também de que genes de resistência ocorrem na forma de famílias gênicas duplicadas (Hammond-Kosack et al., 1997). Assim, produtos não reproduzíveis poderiam ser resultados de eventos aleatórios de hibridização dos iniciadores a alelos destes locos duplicados decorrentes de sua semi-especificidade.

Seqüências reproduzíveis foram encontradas em todas as linhagens e também entre linhagens. Sugere-se que tais seqüências representem alelos dos respectivos locos ESTs, embora esta hipótese só possa ser comprovada através de estudos de co-segregação. Neste caso, no entanto, o número de alelos encontrados (4 a 6) ficou abaixo daquele esperado no caso de uma

espécie altamente poliplóide. Segundo a literatura, o grau da ploidia de cana-de-açúcar é altamente variável tanto entre plantas oriundas de reprodução sexuada como entre cromossomos de uma mesma planta. O número médio estimado de cada cromossomos gira ao redor de 8 a 10 (D'Hont et al., 1993). A discrepância entre o número de seqüências reproduzíveis, que supostamente representam alelos, e o número esperado de alelos com base em estudos citogenéticos, pode plausivelmente ser explicada pela presença de alelos idênticos em alguns cromossomos ou ainda pela presença de alelos que não são amplificados pelos iniciadores utilizados devido à divergência de seqüências nos sítios de hibridização dos mesmos. O primeiro caso explicaria a ocorrência de clones representantes de cada alelo em diferentes freqüências, com alelos duplicados ocorrendo em maior freqüência. No segundo caso, os alelos divergentes poderiam ser eventualmente amplificados em decorrência de variações mínimas não controláveis nas condições da PCR (principalmente na concentração de DNA molde). Estas ampliações eventuais dariam origem a algumas das seqüências não-reproduzíveis.

A comparação das seqüências reproduzíveis revelou a existência de polimorfismos (SNPs) entre seqüências tanto de uma mesma variedade como entre variedades. As seqüências mais polimórficas foram àquelas similares ao gene *Rp1-D* amplificadas na variedade SP11763 (1 SNP/14 pb). Por outro lado, o mesmo fragmento amplificado da variedade R570 foi o menos polimórfico de todos, apresentando apenas 1 polimorfismo a cada 196 pb. Tal fato reveste-se de especial interesse se considerarmos que a primeira variedade é considerada suscetível à ferrugem, ao passo que a segunda é fruto de seleção para resistência (Daugrois et al. 1996). Segue-se que se a reduzida variabilidade nucleotídica na variedade R570 for devida à seleção para resistência a ferrugem, neste caso as seqüências identificadas neste trabalho podem estar ligadas aos genes de resistência selecionados durante o processo ou ainda mesmo serem parte de um loco de resistência.



A variabilidade entre seqüências similares ao gene *Xa1* também variou em função da variedade. No caso da SP80180, a freqüência de SNPs foi de 1/37pb, ao passo que na variedade SP4966 foi o dobro. Estes valores intermediários são similares aos relatados em arroz (1 SNP / 89 pb) (Nasu et al.,2002) e em milho (1 SNP / 70 pb) (Bhatramakki et al.,2000).

A comparação das seqüências reproduzíveis aqui relatadas com aquelas depositadas no banco SUCEST permitiu identificar apenas uma (3/49) idêntica ao EST SCCCCL3080E03.g e similar ao gene *Xa1*. A ausência de outras seqüências idênticas pode ser devida à utilização de uma variedade diferente pelo projeto SUCEST daquelas utilizadas neste trabalho e/ou a falsos polimorfismos devidos à qualidade insatisfatória em algumas regiões das leituras dos ESTs. Merece atenção especial, no entanto, o fato de que SCCCCL3080E03.g é uma uni-seqüência, ou seja, não há nenhuma outra similar a ela em todo o banco SUCEST, e também que este EST foi sequenciado de uma biblioteca construída a partir de calo da variedade SP803280 submetido a estresse de temperatura (4°C e 37°C) e alternância de luz. Neste caso, o padrão de expressão deste EST pode ser igual à do gene *Xa1* de arroz, que é um dos únicos genes de resistência que apresentam expressão induzida, inclusive por injúria mecânica (Yoshimura et al., 1998). Portanto, seria interessante estudar os efeitos fenotípicos da super-expressão do alelo 3/49 em cana, valendo-se da relativa facilidade de transformação e regeneração *in vitro* desta espécie.

No caso das seqüências homólogas ao gene *Rp1-D*, todas foram altamente similares, mas não idênticas, ao cluster 172802, que compreende seis leituras. Destes, dois foram seqüenciados de tecidos expostos a estresse similar ao caso anterior ou ainda infectados com *Acetobacter diazotrophicus*, bactéria endofítica fixadora de nitrogênio.

Com base na ocorrência de polimorfismos, tentou-se desenvolver oligonucleotídeos específicos para determinados alelos das variedades SP80180 e SP804966. A vantagem da utilização de oligonucleotídeos alelo-

específicos, no caso da cana-de-açúcar, reside, ao menos teoricamente, no fato de simplificarem os padrões de segregação observados em populações de mapeamento de espécies polissômicas e poliplóides, justamente por serem específicos para determinados alelos. Neste caso, a escolha de polimorfismos para tal foi grandemente dificultada pela condição polialélica da espécie, uma vez que polimorfismos entre dois alelos foram compensados por monomorfismos entre estes e os demais. Em adição a isto, há que ser lembrado que estes locos, além de multi-alélicos, podem ainda estar duplicados. Assim, oligonucleotídeos podem ser específicos para alelos de determinado loco, mas não necessariamente para alelos de um outro loco homólogo.

Do exposto acima, depreende-se que, a despeito da relativa abundância de polimorfismos de seqüência nucleotídica em cana-de-açúcar e corroborando com Grivet et al. (2001), a utilização de SNPs como marcadores moleculares em cana-de-açúcar baseados em iniciadores alelo-específicos mostra-se um pouco mais complicada do que aquela preconizada para organismos diplóides.

## 6 CONCLUSÕES

1. A comparação das seqüências reproduzíveis revelou a existência de polimorfismos (SNPs) entre seqüências tanto de uma mesma variedade como entre variedades;
2. O número de alelos encontrados (4 a 6) ficou abaixo daquele esperado no caso de uma espécie altamente poliplóide;
3. As seqüências mais polimórficas foram àquelas similares ao gene *Rp1-D* amplificadas na variedade SP11763 (1 SNP/14 pb). O mesmo fragmento amplificado da variedade R570 foi o menos polimórfico de todos, apresentando apenas 1 polimorfismo a cada 196 pb.
4. Na variedade SP80180, a freqüência de SNPs foi de 1/37pb, ao passo que na variedade SP4966 foi o dobro 1/75 pb;
5. A utilização de SNPs como marcadores moleculares em cana-de-açúcar baseado em iniciadores alelo-específicos foi infrutífera, provavelmente devido a condição polialélica da espécie estudada.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALTSCHUL, S.F.; MADDEN, T.L.; SHAFFER, A.A.; ZHANG, Z.; MILLER, W.; LIPMAN, D.J. “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”. **Nucleic Acids Research**, v.25, p.3389-3402, 1997.
- THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature**, v.408, p.796–815, 2000.
- ARRUDA, P. Sugarcane transcriptome. A landmark in plant genomics in the tropics. **Genetics and Molecular Biology**, v.24, p.3, 2001.
- BENT, A.F. Plant disease resistance genes: function meets structure. **Plant Cell**, v.8, p.1757-1771, 1996.
- BHATRAMAKKI, D.; CHING, A.; MORGANTE, M.; DOLAN, M.; REGISTER, J.; SMITH, H.; TINGEY, S.; RAFALSKI, J.A. Conserved single nucleotide polymorphism (SNP) haplotypes in maize. In: PLANT AND ANIMAL GENOME CONFERENCE, 8., San Diego, 2000. **Abstracts**. San Diego: Scherago International, 2000. p. 188.

- CAMARGO, L.E.A. Disease resistance related ESTs of sugarcane. In: PLANT AND ANIMAL GENOME CONFERENCE, 8., San Diego, 2000. **Abstracts**. San Diego: Scherago International, 2000. p. 30.
- CARNEIRO, N.P.; CARNEIRO, A.A.; GUIMARÃES, C.T.; PAIVA, E. Desvendando o código genético. **Biotecnologia Ciência & Desenvolvimento**, n.17, p.50-58, 2000.
- COLLINS, A.; LONJOU, C.; MORTON, N.E. Genetic epidemiology of single-nucleotide polymorphisms. **Proceedings of the National Academy of Sciences of the United States of America**, v.96, p.15173-15177. 1999.
- COLLINS, N.C.; JEFF, D.; AYLIFFE, M.; ELLIS, J.; HULBERT, S.; PRYOR, T. Molecular characterization of the maize Rp1-D rust resistance haplotype and its mutants. **The Plant Cell**, v.11, p.1365-1376, 1999.
- COLLINS, N.C.; WEBB, C.A.; SEAH, S.; ELLIS, J.G.; HULBERT, S.H.; PRYOR, A. The isolation and mapping of disease resistance gene analogs in maize. **Molecular Plant-Microbe Interactions**, v.11, p.968-978, 1998.
- DAUGROIS, J.H; GRIVET, L.; ROQUES, D.; HOARAU, J.Y.; LOMBARD, H.; GLASZMANN, J.C.; D'HONT, A. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570". **Theoretical and Applied Genetics**, v.92 p.1059–1064, 1996.
- D'HONT, A.; LU, Y.H.; LÉON, D.G.; GRIVET, L.; FELDMANN, P.; LANAUD, C.; GLASZMANN, J.C. A molecular approach to unraveling the genetics of

sugarcane, a complex polyploid of the Andropogoneae tribe. **Genome**, v.37, p.222-230, 1993.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO.

**Sugarcane Expressed Sequence Tag (SUCESTE).**

<http://sucest.lad.ic.unicamp.br> (18 jun. 2002).

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO.

**Organization for Nucleotide Sequencing and Analysis (ONSA).**

<http://watson.fapesp.br> (22 Jun. 2002).

GRIVET, L.; GLASZMANN, J.C.; ARRUDA, P. Sequence polymorphism from EST data in sugarcane: a fine analysis of 6-phosphogluconate dehydrogenase genes. **Genetics and Molecular Biology**, v.24 p.161–167, 2001.

GRIVET, L.; D'HONT, A.; DUFOUR, P.; HAMON, P.; ROQUES, D.; GLASZMANN, J.C. Comparative genome mapping of sugar cane with other species within the Andropogonae tribe. **Heredity**, v.73, p.500-508, 1994.

GRIVET, L.; D'HONT, A.; ROQUES, D.; FELDMANN, P.; LANAUD, C.; GLASZMANN, J.C. RFLP Mapping in cultivated sugarcane (*Sacharum spp.*): genome organization in highly polyploid and aneuploid interspecific hybrid. **Genetics**, v.142, p.987-1000, 1996.

HAMMER, M. F.; SPURDLE A. B.; KARAFET T.; BONNER M. R.; WOOD E. T.; NOVELLETTO A.; MALASPINA, P.; MITCHELL R. J.; HORAI S.; JENKINS

- T.; ZEGURA S. L. The geographic distribution of human Y chromosome variation. **Genetics**, v.145, p.787-805,1997.
- HAMMOND-KOSACK, K.E., JONES, J.D.G. Plant disease resistance genes. **Annual Review of Plant Physiology and Plant Molecular Biology**, v.48, p.575-607, 1997.
- HANAHAN, D. Studies on transformation of *Escherichia coli* with plasmids. **Journal of Molecular Biology**, v.166, n. 4, p. 557-580,1983.
- HOWARD, T.D.; BLEECKER, E.R.; STINE, C.O. Fluorescent allele-specific PCR (FAS-PCR) improves the reliability of single nucleotide polymorphism screenig. **Bio. Techniques**, v.26, p.380-381, 1999.
- HULBERT, S.H. Structure and evolution of the *rp1* complex conferring rust resistance in maize. **Annual Review of Phytopathology**, v.35, p.293-310, 1997.
- HULBERT, S.H.; WEBB, C.A.; SMITH, S.M.; SUN, Q. Resistance gene complexes: evolution and utilization. **Annual Review of Phytopathology**, v.39, p.285-312, 2001.
- JOHAL. G.S.; BRIGGS, S.P. Reductase activity encoded by the *HM1* disease resistance gene in maize. **Science**, v.258, p.985-987, 1992.
- KANAZIN, V.; MAREK, L.F.; SCHOEMAKER, R.C. Resistance gene analogs are conserved and clustered in soybean. **Proceedings of the National Academy of Sciences of the United States of America**, v.93, p.11746-11750, 1996.

KUNKEL, T.A.; BEBENEK, K. DNA replication fidelity **Annual Review of Biochemistry**, v.69 p.497-529, 2000.

LEISTER, D.; BALLVORA, A.; SALAMINI, F.; GEBHART, C. A PCR-based approach for isolating pathogen resistance genes from potato with potencial for wide application in plants. **Nature Genetics**, v.14, p.421-429, 1996.

LEWIN, B. **Genes VII**. Porto Alegre: Artmed, 2001. p.46-47: De genes a Genomas.

LU, Y.H.; D'HONT, A.; PAULET, F.; GRIVET, L.; ARNAUD, M.; GLASZMANN, J.C. Molecular diversity and genome structure in modern sugarcane varieties. **Euphytica**, v.78, p.217-226, 1994.

MARTIN, G.B.; BROMMONSCHENKEL, S.H.; CHUNWONGSE, J.; FRARY, A.; GANAL, M.W.; SPIVEY, R.; WU, T.; EARLE, E.D.; TANKSLEY, S.D. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. **Science**, v.262, p.1432-1436, 1993.

MATSUOKA, S.; GARCIA, A.A.F.; CALHEIROS, G.G. Hibridação em cana-de-açúcar. In: BOREM, A. **Hibridação artificial de plantas**. Viçosa: UFV, 1999. p.221-254.

MELOTTO, M.; KELLY, J.D. Fine mapping of the Co-4 locus of common bean reveals a resistance gene candidate, COK-4, that encodes for a protein kinase. **Theoretical and Applied Genetics**, v.103, p.508-517, 2001.

MOORE, G.; DEVOS, K.M.; WANG, Z.; GALE, M.D. Grasses line up and form a circle. **Current Biology**, v.5, 737-738, 1995.



- NASU, S.; SUZUKI, J.; OHTA, R.; HASEGAWA, K.; YUI, R.; KITAZAWA, N.; MONNA, L.; MINOBE, Y. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. **DNA Research**, v. 9, p. 163-171, 2002.
- NEWTON, C.A.; GRAHAM, G.A. **PCR**. 2.ed. New York: Springer-Verlag, 1997. 192p.
- NICKERSON, D.A.; TAYLOR, S.L.; WEISS, K.M.; CLARK, A.G.; HUTCHINSON, R.G.; STENGARD, J.; SALOMAA, V.; VARTIANIEN, E.; BOERWINKLE, E.; SING, C.F. DNA sequence diversity in a 9.7 kb region of the human lipoprotein lipase gene. **Nature Genetics**, v.19, p.233-240, 1998.
- OLIVEIRA, M.; VASCONCELOS, Y. Soluções para geração de eletricidade. **Pesquisa FAPESP**, v.65, p. 66-69, 2001.
- PATERSON, A.H.; LIU, S.; LIN, Y.; CHANG, C. **Genome mapping in plants: status of genome , mapping in the tropical grains**. New York: Academic Press, 1996. 330p.
- PENNISI, E. A closer look at SNPs suggests difficulties. **Science**, v.281, p.1787-1789, 1998.
- PEREIRA, M.G.; LEE, M. Identification of genomic regions affecting plant height in sorghum and maize. **Theoretical and Applied Genetics**, v.90, p.380-388, 1995.

- SANGER, F.; NICKLEN, S.; COULSON, R.E. DNA sequencing with chain terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, v.74, p.5463-5467, 1977.
- SANTOS, F.R. PCR haploty for the humany chromossome based on alphaid satellite DNA variants and heteroduplex analysis. **Genes**, v.165 p.199-208, 1995,
- SILVA, J.A.G.; SORRELLS, M.E.; BURNQUIST, W.L.; TANKSLEY, S.D. RFLP linkage map and genome analysis of *Saccharum spontaneum*. **Genome**, v.36 p.782-791, 1993.
- SOMERS, D.J.; KIRKPATRICK, R.; MONIWA, M. Characterization of single nucleotide polymorphisms in hexaploid wheat. In: PLANT, ANIMAL AND MICROBE GENOME CONFERENCE, 10., San Diego, 2002. **Abstracts**. San Diego: Scherago International, 2002. p.193.
- SOORI, V.A.; WATANABE, K.N.; VALKONEN, J.P.T. Predicted kinase-3a motif of a resistance gene analogue as a unique marker for virus resistance. **Theoretical and Applied Genetics**, v.99, p.164-170,1999.
- SUN, Q.; COLLINS, C.N.; AYLIFFE, M.; SMITH, S.M.; DRAKE, J.; PRYOR, T.; HULBERT, S.H. Recombination between paralogues at the *rp1* rust resistance locus in maize. **Genetics**, v. 158, p. 423-438, 2001.
- VETTORE, A.L.; SILVA, F.R.; KEMPER, E.L.; ARRUDA, P. The libraries that made SUCEST. **Genetics and Molecular Biology**, v.24, p.1-4, 2001.

- WOLTERS, P.; POWELL, W.; LAGUDAH, E.; SNAPE, J.; HENDERSON, K.  
Nucleotide diversity at homoeologous loci in wheat. In: PLANT AND ANIMAL GENOME CONFERENCE, 8., San Diego, 2000. **Abstracts**. San Diego: Scherago International, 2000. p. 103.
- YOSHIMURA, S.; YAMANOUCHI, U.; KATAYOSE, Y.; TOKI, S.; WANG, I.K.; KURATA, N.; IWATA, N. SASAKI, T. Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. **Proceedings of the National Academy of Sciences of the United States of America**, v.95, n.4, p.1663-1668, 1998.
- YU, Y.G.; SAGHAL, M.A.; ROOF, M.A. Isolation of a superfamily of candidate disease – resistance genes in soybean based on a conserved nucleotide – binding site. **Proceedings of the National Academy of Sciences of the United States of America**, v.95, p.1663-1668, 1998.